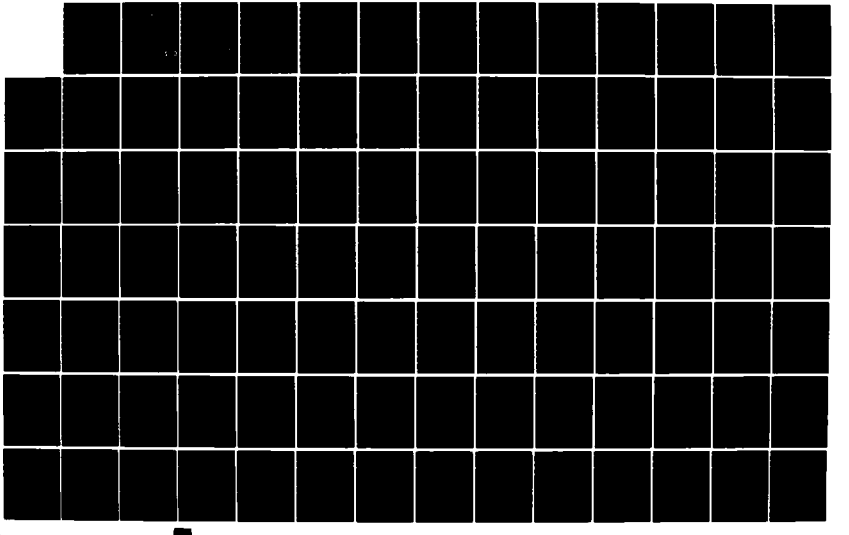
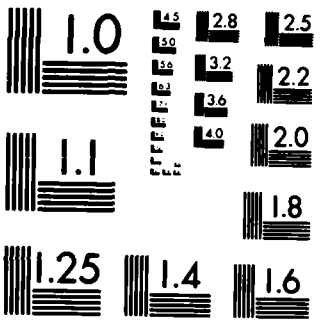
AD-A151 853 A NON-PARAMETRIC PROBABILITY DENSITY ESTIMATOR AND SOME APPLICATIONS(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF ENGINEERING R P FUCHS MAY 84 1/2

UNCLASSIFIED AFIT/DS/ENC/84-1 F/G 12/1 NL





A NON-PARAMETRIC PROBABILITY DENSITY ESTIMATOR AND SOME APPLICATIONS

DISSERTATION

Ronald P. Fuchs, B.S., M.S.
Major, USAF

AFIT/DS/ENC/84-1

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY
AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT/DS/ENC/84-1

A NON-PARAMETRIC PROBABILITY DENSITY ESTIMATOR AND SOME APPLICATIONS

DISSERTATION

Ronald P. Fuchs, B.S., M.S.
Major, USAF

AFIT/DS/ENC/84-1

Approved for public release; distribution unlimited

AFIT/DS/ENC/84-1

A NON-PARAMETRIC PROBABILITY DENSITY ESTIMATOR AND SOME APPLICATIONS

DISSERTATION

Presented to the Faculty of the School of Engineering
of the Air Force Institute of Technology
Air University
In Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy

Ronald P. Fuchs, B.S., M.S.
Major, USAF

May 1984

AFIT/DS/ENC/84-1

A NON-PARAMETRIC PROBABILITY DENSITY ESTIMATOR AND SOME APPLICATIONS

Ronald P. Fuchs, B.S., M.S.

Major, USAF

Approved:

Gilbert H Moore — 16 May 1984

Joseph P. Ca — 18 May 1984

[illegible] — 15 May 1984

Joseph J. Coleman — 11 May 1984

Accepted:

J. P. Bennington 18 May 1984

Dean, School of Engineering

Preface

The work presented in this dissertation was inspired by suggested improvements to a density estimation technique developed by Major Jim Sweeder. My committee chairman, Dr. Albert H. Moore, was responsible for these key ideas and many other helpful suggestions during the course of this research. He, along with every faculty member with whom I have dealt in my program at the Air Force Institute of Technology (AFIT), has shown enthusiastic desire for each student to learn and the willingness to help in this task. This attitude has made my stay at AFIT quite pleasant and I am grateful.

The F-16 Systems Program Office sponsored this research and provided support particularly in the area of computer resources. I thank them and hope they can use the results of this study.

My student colleagues have often been helpful with advice and criticism. I particularly appreciate the efforts of Major Max Stafford to keep me on a sound mathematical basis, and those of Captain Ron Hinrichsen to facilitate the implementation of some rather extensive computer programs.

I could never have completed this task without the loving support of my wife, Sally. There were many difficult periods in the program but she always made them easier. My children, Alison and Adam, did not always make

things easier but did keep my efforts in perspective. I love and appreciate all of them.

In many cases, the reference cited are not inclusive, but are prominent in the field or contain extensive bibliographies. The Bibliography in this dissertation is more detailed than the text references and is intended to provide a good foundation for those wishing to further research the field of non-parametric density estimation.

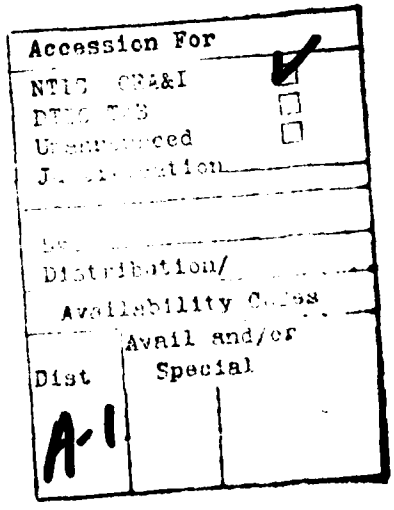


Table of Contents

List of Figures

List of Tables

Abstract

A new non-parametric probability density estimator is developed which has the following properties:

1) It yields a continuous, non-negative and piecewise linear estimate.

2) It converges to the true density function if the true density has no more than a finite number of discontinuities of a form where the value of the function at the discontinuity can be considered the average of the limiting values on either side of the discontinuity.

3) It requires no user supplied parameters.

The estimator is shown to have significantly better error properties, for certain classes of distributions, than existing density estimators. The quality of the estimate is discussed, tabulated and graphically demonstrated. Applications, including parameterization, small sample analysis, and two sample tests are presented. These newly developed applications are shown to improve upon the generally accepted existing techniques. Guidelines for choosing a density estimation method along with an organized approach to method selection are discussed.

I. Introduction.

The historical development of non-parametric probability density function estimators stems from the histogram type estimator which was inspired by John Graunt and further developed by mathematicians such as Petty, Huygens, van Dael and Halley (230). Density estimation has been attempted by distinguished statisticians including Pearson, Gossett, Fisher, Johnson, et.al.(52,143,203). Their methods include methods of parameterization, kernel estimators, distance estimation, entropy methods, spline techniques and series estimators. This dissertation presents a new non-parametric density estimator.

A question which is logically addressed is: "What good is a density estimator?" Some uses of density estimators were discussed by Sweeder (202) and much of the work presented in this dissertation is an extension of his groundbreaking efforts. Some other uses of density estimates are discussed throughout this paper. The specific applications presented by Sweeder were avoided here since redoing them with a slightly different estimate seemed rather trivial. Some new applications of density estimates are presented in this dissertation. In particular, a two-sample test is discussed which takes advantage of the potentially large difference created by an unbounded operator acting upon relatively small differences in the

CDFs. The intent of this dissertation is to develop the actual estimator and show the use of an estimator of this type. Many of the "proofs" rely on empirical evidence obtained from tremendously expensive Monte Carlo analysis. In these cases only enough of the Monte Carlo runs were completed to demonstrate the techniques and results.

Throughout this dissertation, comparisons will be made among results from samples from uniform, normal, and double exponential (Laplace) distributions. The estimator developed is not limited to these, or even symmetric, distributions, but for comparison purposes with previous research (226) much of the work presented here uses these three distributions, which are assumed to be representative of platykurtic, mesokurtic, and leptokurtic distributions in general.

The dissertation is divided into four main sections (Chapters II-V). The first discusses the development of the estimator itself, the underlying theory, and the trade-offs made in its development along with the reasons for those trade-offs. The second main section is essentially a validation of the estimator developed in the first section. Both graphical and tabular comparisons of results are given. The third section presents some applications including parameterization through distance estimation, a new small sample analysis technique, and a new two-sample test. Other possible applications are dis-

cussed. The last main section was inspired by a goal which was set during the definition phase of this research program. That goal was to develop a density estimator which could be used by the relatively uninitiated without the requirement to choose any parameters. This section presents some general, easy to understand and apply guidelines for when to use this, or for that matter any, nonparametric estimator. Supporting data for a choice between this estimator and some others is presented.

The final chapter summarizes the results of this research effort. There is always another step to be taken in research and Chapter VI discusses several possible directions in which to take that step. Hopefully it will be of use to those continuing down the path to better nonparametric density estimators and new applications of those estimators.

II. The Estimator.

Non-parametric density function estimators suffer, to one extent or another, from some or all of the following problems:

1) They require user specified "parameters" which can greatly affect the shape of the estimated function, but cannot be, or are not easily, optimally determined. This problem is exacerbated when the estimator is overly sensitive to these "parameters". For example, the maximum penalized likelihood estimator (39,178,203) requires two such parameters. Although it is theoretically possible to find the optimal values, realistically the values are determined by trial and error. This makes density estimation an art, with the result that, when this particular estimator is used by the unskilled, all estimates tend to look like normal density functions. Since this estimator and a kernel estimator with similar problems are the only ones commonly available (they appear in the International Mathematical and Statistical Libraries (IMSL) package of FORTRAN subroutines available through IMSL, 7500 Bellaire Blvd., Houston, TX, 77036), many potential users may have rejected non-parametric density estimation as too difficult or not accurate enough.

2) They tend to be noisy, like the frequency polygon estimator (203). This can be corrected by averaging or

other smoothing processes such as Sweeder used. Many of the Bootstrap techniques (47) are suitable for this job. Frequency domain smoothing via Fourier transform analysis may also be used.

3) They are not uniquely defined, particularly for small samples. That is, one may obtain an entirely different estimate by slightly varying a parameter of the estimator. The histogram estimator typifies this problem.

4) They only give reasonable estimates for relatively large samples. This is a problem in virtually all nonparametric estimators (Sweeder's being a notable exception.) Unfortunately, in many cases, large samples are difficult or expensive to obtain.

5) They require restrictive assumptions about the form of the underlying distribution (i.e. symmetry, unimodality, infinite or finite support, etc.)

6) They do not balance sensitivity and robustness. That is, they tend to either give the same density shape for samples from a wide variety of distributions, or they are overly sensitive to sample peculiarities such as outliers or closely grouped data points. The very nature of a random sample makes these deficiencies difficult to handle. For example, if one makes adjustments to the estimator to take into account close spacing of sample points, then true peaks in the density will be rounded and true valleys will be filled.

7) They result in an infeasible estimate. Many common density estimators yield negative densities, others estimate support which does not include the entire sample.

The above problems cannot all be solved simultaneously. The interactions among these areas is what makes density estimation so difficult.

All non-parametric density estimators have the additional problem of estimating the support for the density. This is usually handled in one of the following manners:

1) Estimate $f(x \mid x_{(1)} \leq x \leq x_{(n)})$

2) Estimate the support based on some sample extrapolation rule.

3) Assume some support based on knowledge of the data source, for example $(0, \infty)$, $(0,1)$, $(-\infty, \infty)$, etc. This is a sort of Bayesian non-parametric estimation.

4) Estimate the support from the extreme order statistics from a set of samples. That is, estimate the distribution of $x_{(1)}$ and $x_{(n)}$ and select some percentage point of these distributions as the estimate of the endpoint (64). There is seldom enough data available to actually use this method.

Endpoint estimation techniques used in this estimator will be discussed later in this chapter. For now we assume that the density is non-zero only on the interval $[x_{(0)}, x_{(n+1)}]$, that values of $x_{(0)}$ and $x_{(n+1)}$ have already been defined or estimated, and that these values converge

to the true support of the distribution as the sample size increases.

II.1 *Development of the Estimator*

Consider a random sample, $x_1, x_2, x_3, \ldots, x_n$, of size n from an unknown univariate, continuous probability distribution function, F(x). Let $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$ represent the ordered random sample such that $x_{(1)} \leq x_{(2)} \leq \ldots \leq x_{(n)}$. Now define $G_i = G(x_{(i)})$, $i=1,2,\ldots,n$ be the plotting rule that is associated with the i^{th} order statistic. G_i is a value of the sample distribution function at this point, of the form $G_i = (i+\alpha)/(n+\beta)$, with $-1 \leq \alpha \leq \beta \leq 1$ (we will discuss selection of plotting rule parameters α and β in more detail later in this chapter.) Let

$$\Delta G = G_i - G_{i-1} = 1/(n+\beta)$$

We know that

$$\int_{x_{(i-1)}}^{x_{(i)}} f(x)dx = F(x_{(i)}) - F(x_{(i-1)})$$

if we approximate

$$F(x_{(i)}) - F(x_{(i-1)}) = \Delta G$$

and assume that f(x) varies linearly between $x_{(i-1)}$ and

$x_{(i)}$ we obtain:

$$\hat{f}_i = 2\Delta G/(x_{(i)} - x_{(i-1)}) - \hat{f}_{i-1}$$

where

$$\hat{f}_i = \hat{f}(x_{(i)})$$

For a plotting rule using $\beta = 0$ this is similar to the classical frequency polygon estimator.

This estimator has some nasty properties. The value at some points may be negative since $\hat{f}_{i-1}$ is not guaranteed to be less than $(2\Delta G)/(x_{(i)}-x_{(i-1)})$. In addition, since $\hat{f}(x_{(0)})$ or $\hat{f}(x_{(n+1)})$ may be arbitrarily defined there are an infinite number of possible estimators. Even if we define the density as zero at the endpoints the estimation process can be started at either end and the result will, in general, depend upon the end at which we start. This means that the estimator is dependent upon the path taken through the sample.

Both of these undesirable characteristics may be corrected. Assume some $\hat{f}_i$ is the first estimate calculated as a negative value. Let $\hat{f}_i^* = 0$,
and set

$$\hat{f}_{i-1}^* = [4\Delta G - \hat{f}_{i-1}(x_{(i-1)} - x_{(i-2)})]/(x_{(i)} - x_{(i-2)})$$

The next calculated value, $\hat{f}_{i+1}^* = 0$, will always be greater than zero, as will $\hat{f}_{i-1}^* = 0$ (See Figure 1). The

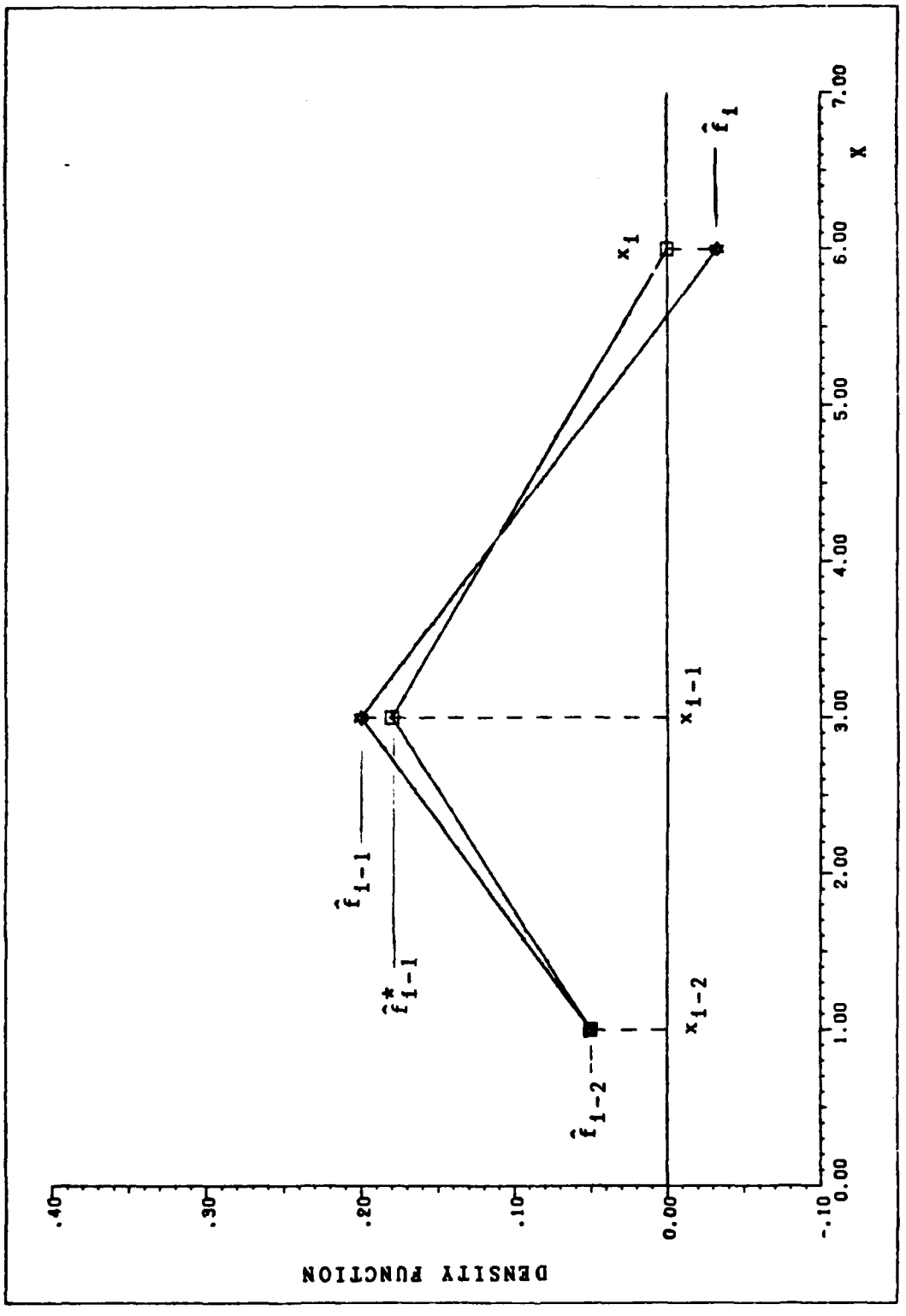


Figure 1 - Eliminating Negative Density Estimates

result of this process is a piecewise linear, non-negative estimate of f(x) given by:

$$\hat{f}(x) = \hat{f}_{i-1} + (x-x_{(i-1)})(\hat{f}_i-\hat{f}_{i-1})/(x_{(i)}-x_{(i-1)})$$

$$x_{(i-1)} \leq x \leq x_{(i)}$$

and

$$\hat{f}(x) = 0 \qquad x \notin [x_{(0)}, x_{(n+1)}]$$

Which, when integrated, yields a continuous, piecewise quadratic distribution function, $\hat{F}(x)$.

In order to remove the ambiguity in $\hat{f}(x)$ which exists from the possibility of starting the process at either end, we calculate the forward estimate, $\bar{\hat{f}}(x)$, and the backward estimate, $\underline{\hat{f}}(x)$, and average the two to obtain $\hat{f}(x)$. This process not only removes the path ambiguity but also tends to eliminate zero values of the density estimate introduced in order to assure non-negativity of $\hat{f}(x)$. Figures 2(a) and 2(b) show the results of using the estimator, as described so far, on random samples of size 100 from two distributions. Notice that this estimator is quite noisy. We will consider a solution to this problem shortly.

The estimator does have some desirable properties when we consider the distribution function estimate.

1) $\hat{F}(x)$ is differentiable everywhere.
2) $\hat{F}(x)$ is a distribution function.
3) $G_{i-1} \leq \hat{F}(x_{(i)}) \leq G_i \qquad i=1,2,\ldots,n$

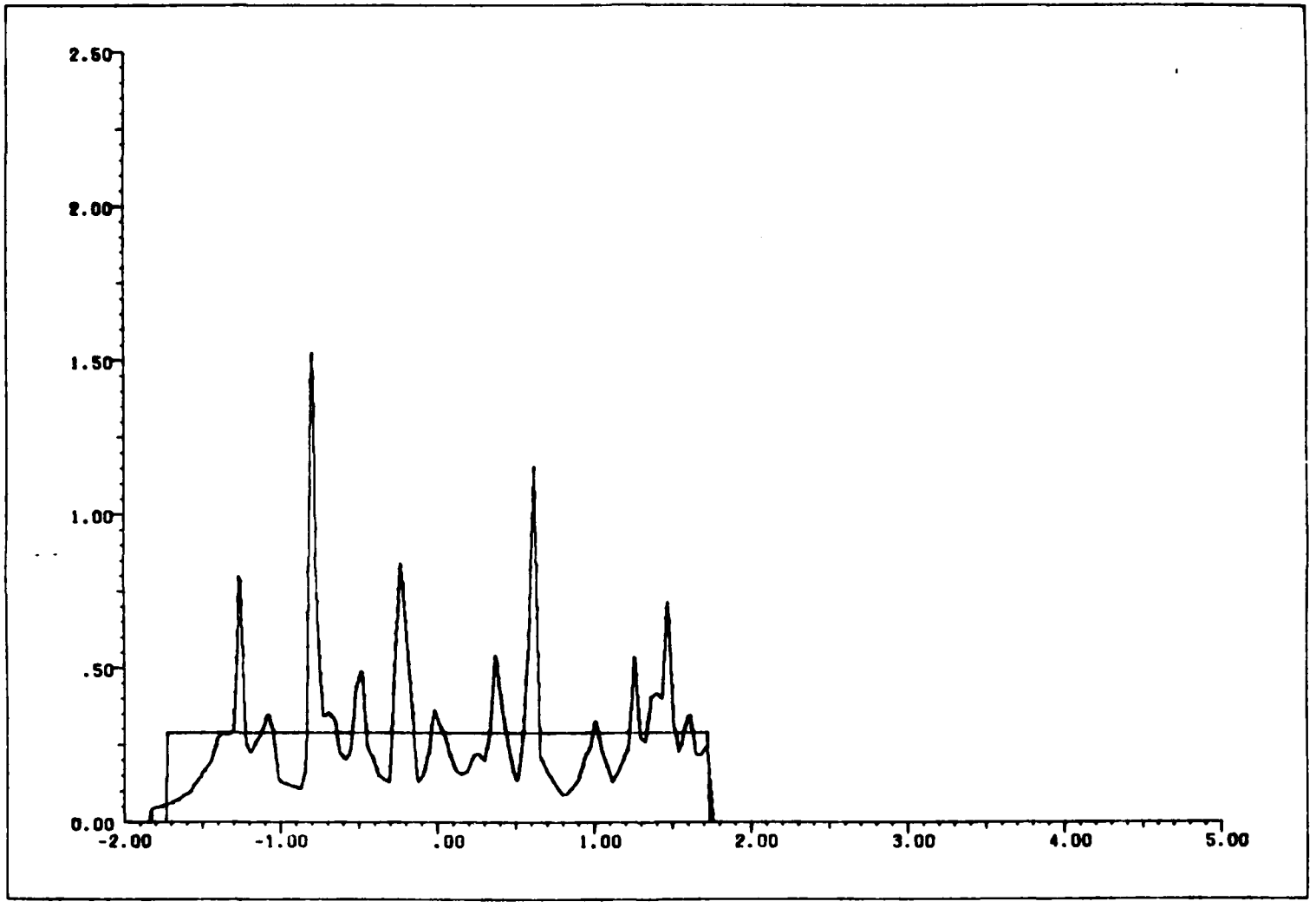


Figure 2(a) - Raw Density Estimate for Uniform Sample (n=100)

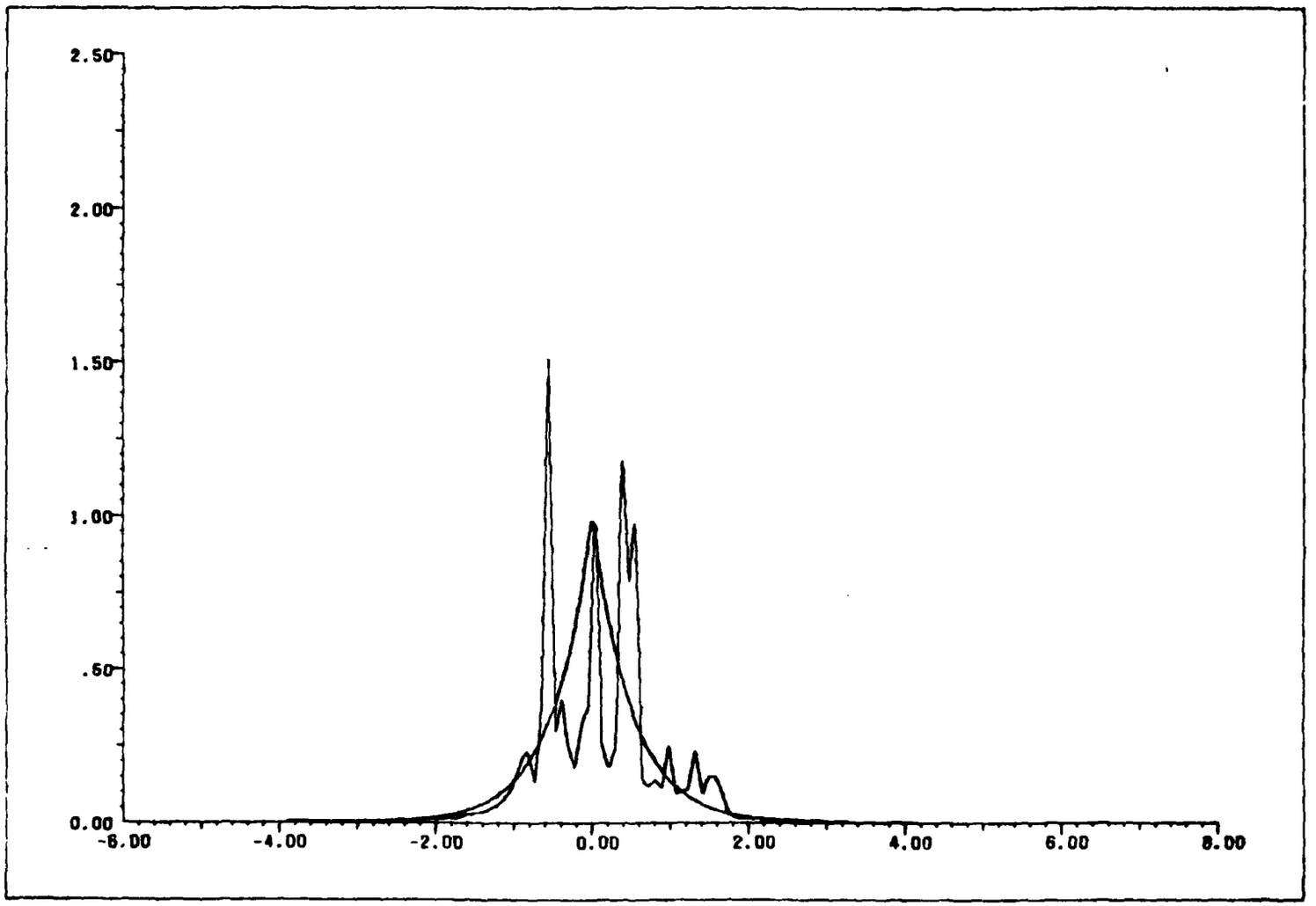


Figure 2(b) - Raw Density Estimate for Laplace Sample (n=100)

These properties are essentially the same as those of Sweeder's estimator. However this estimator has the additional property that it does not necessarily go to zero at the sample points. It is this feature that will allow us to show convergence to the true density, and reduce the amount of smoothing required to obtain a "good" density estimate.

Analogous to Sweeder we may use a Bootstrap (47) type technique to obtain some smoothing of our estimator. This is desirable for cases where we have unnaturally closely grouped data within the sample (i.e. data which does not reflect the true character of the underlying distribution.) Experimentation with samples from known distributions indicates that this is a problem which occurs frequently in small sample situations. We choose d subsamples from our original sample as follows:

$$\{x_{(j)};\ j=k+md,\ k=1,2,\ldots,d;\ d<\infty\ ;$$
$$d\leq n/2;\ m=0,1,2,\ldots,[(n-k)/d]\}$$

Estimates are calculated using each of these subsamples successively so that we have estimates, $\hat{f}_j(x)$, $j=1,2,\ldots,d$. The estimator for $f(x)$ is obtained by averaging:

$$\hat{f}(x) = \frac{1}{d}\sum_{j-1}^{d}\hat{f}_j(x)$$

II.2 Properties of the Estimator

A desirable property of an estimator is that it converges to the true function as the sample size increases. The estimator obtained so far has this property as will be shown in the following theorems, but first some fundamental definitions.

Let R be the real line, B a Borel field on R and P a probability measure defined on B. The function F defined on (R,B,P) by $F(x) = P(\{I:I=(-\infty,x]\ R\})$ is the distribution function of P.

$\hat{F}_n(x)$, as we have defined it, is a distribution function, since:

1) $\hat{F}_n(x) = \int_{-\infty}^{x} \hat{f}_n(x)dx$ is non-decreasing

since $\hat{f}_n(x)$ is non-negative by construction.

2) $\hat{F}_n(x)$ is continuous by construction

3) $\hat{F}_n(x) = 0 \quad x < x_{(0)}$

$\hat{F}_n(x) = 1 \quad x > x_{(n+1)}$ by definition

Therefore $\hat{F}_n(x)$ is a probability distribution function.

The following development also assumes a random sample, $x_1,x_2,\ldots,x_n$ from a continuous distribution, F(x), and that $F'(x)$ exists. Parenthetical subscripts are again used to represent the ordered sample.

Lemma 1 - A finite convex combination of sequences of functions, each of which converges (uniformly) to a single function, will converge (uniformly) to that function.

Proof - Let $\{f_{ij}(x)\}_{j=1,2,3,\ldots}^{i=1,2,\ldots,k}$ be a set of sequences with

$$|f_{ij}(x) - f(x)| < \epsilon \qquad j > N \qquad i=1,2,\ldots,k$$

and let

$$S(x) = \sum_{i=1}^{k} \alpha_i f_{ij}(x) \; ; \quad \sum_{i-1}^{k} \alpha_i = 1 \; ; \quad \alpha_i \geq 0$$

then

$$|S(x) - \sum \alpha_i f(x)| = |\sum \alpha_i f_{ij}(x) - \sum \alpha_i f(x)| =$$

$$|\sum \alpha_i (f_{ij}(x) - f(x))| \leq \sum \alpha_i |f_{ij}(x) - f(x)| <$$

$$\sum \alpha_i \epsilon = \epsilon$$

The extension to uniform convergence is analogous if we start with the hypothesis:

$$|f_{ij}(x) - f(x)| < \epsilon \; ; \qquad j > N \; ; \quad \forall \; x \in \mathbb{R}$$

Lemma 2 - Given a partition, $P_n = \{x_1 \leq x_2 \leq \ldots \leq x_n\}$ of (a,b) such that $g(x;P_n) \longrightarrow g(x)$, any evenly divided sub-partition,

$$\tilde{P}_m = \{x_{(j)};\ j=k+md,\ k=1,2,\ldots,d;\ d<\infty\ ;$$
$$d\leq n/2;\ m=0,1,2,\ldots,[(n-k)/d]\}$$

results in $g(x;\tilde{P}_m) \xrightarrow{\text{unif}} g(x)$.

Proof -

$$|g(x;P_p) - g(x;P_q)| \leq \epsilon/2\ ;\ \forall\ x \in (a,b)\ ;\ p,q > N_1 \qquad (1)$$

Since P_p and P_q are partitions and we have uniform convergence, this is the Cauchy condition. Also

$$|g(x) - g(x;P_p)| \leq \epsilon/2\ ;\quad x \in (a,b)\ ;\ p > N_2$$
by definition.

Now let $N = d\ \max(N_1,N_2)$ then from (1) we have

$$|g(x;P_p) - g(x;\tilde{P}_N)| \leq \epsilon/2$$

and

$$|g(x;\tilde{P}_N) - g(x)| \leq |g(x;P_p) - g(x;\tilde{P}_N)| +$$
$$|g(x) - g(x;P_p)| \leq \epsilon$$

so

$$|g(x;\tilde{P}_m) - g(x)| \leq \epsilon \quad x \in (a,b) \quad m \geq N$$

That is:

$$g(x;\tilde{P}_m) \xrightarrow{\text{unif}} g(x)$$

The above Lemmas allow us to prove convergence of the basic estimator and then extend it easily to the subsampled case without directly considering each estimator based on a subsample.

Theorem 1 - The sequences $\hat{F}_j(x)$ converge almost everywhere to F(x) where:

$$\hat{F}_j(x) = \begin{cases} 0 & x \le x_{j(0)} \\ \frac{i-1+\alpha}{n+\beta} + \hat{f}_{j(i-1)}(x-x_{j(i-1)}) + (\hat{f}_{j(i)}-\hat{f}_{j(i-1)})(x-x_{j(i-1)})^2 / (x_{j(i)}-x_{j(i-1)}) & x_{j(i-1)} \le x \le x_{j(i)} \\ 1 & x \ge x_{j(m+1)} \end{cases}$$

and

$$\hat{f}_j(x) = \begin{cases} \frac{2}{(m+\beta)(x_{j(i)}-x_{j(i-1)})} - \hat{f}_{j(i-1)} & x_{j(i-1)} \le x \le x_{j(i)} \\ 0 & \text{otherwise} \end{cases}$$

(Since j represents the index of a sequence based upon a particular subsample, and we intend to show the proof for one subsample and later extend it using Lemmas 1 and 2, we will temporarily drop the j subscript for simplicity.)

Proof - Consider the points $x_{(i)}$; $i = 1,2,\ldots,n+1$

$$\hat{F}(x_{(i)}) = \begin{cases} 0 & i=0 \\ G_i & i=1,2,\ldots,n \\ 1 & i=n+1 \end{cases}$$

This is essentially the empirical distribution function, E(x), which has been shown to converge almost everywhere to F(x).

That is

$$\lim_{n\to\infty} E_n(x_{(i)}) = \lim_{n\to\infty} \hat{F}(x_{(i)}) = \lim_{n\to\infty} G_i = F(x_{(i)})$$

For any x in the interval $[x_{(i-1)},x_{(i)})$ we know:

$$\hat{F}(x_{(i-1)}) \leq \hat{F}(x) \leq \hat{F}(x_{(i)})$$

from the monotone property of F(x).

So:

$$\lim \hat{F}(x_{(i-1)}) \leq \lim \hat{F}(x) \leq \lim \hat{F}(x_{(i)})$$

(where we denote $\lim_{n\to\infty}$ as lim) or:

$$\lim G_{i-1} \leq \lim \hat{F}(x) \leq \lim G_i$$

$$F(x_{(i-1)}) \leq \lim \hat{F}(x) \leq F(x_{(i)})$$

and since

$$F(x_{(i-1)}) \leq F(x) \leq F(x_{(i)})$$

$$|F(x) - \lim \hat{F}(x)| \leq F(x_{(i)}) - F(x_{(i-1)}) =$$

$$\lim \frac{i+\alpha}{n+\beta} - \lim \frac{i-1+\alpha}{n+\beta} = \lim \frac{1}{n+\beta} = 0$$

or

$$\lim_{n\to\infty} \hat{F}_j(x) = F(x) \qquad \text{almost everywhere (a.e.)}$$

and by Lemma 1 we have:

$$\hat{F}(x) = \frac{1}{d} \sum_{j=1}^{d} \hat{F}_j(x) \longrightarrow F(x) \qquad \text{a.e.}$$

We have shown convergence of the distribution function. We now proceed to show the more powerful result, convergence of the density function estimate to the true density function.

Theorem 2 - The density function estimate, $\hat{f}(x)$, converges almost everywhere to the true density function, $F'(x) = f(x)$, provided $f(x)$ exists and is continuous.

Proof - For some point x in the interval $[x_{(i-1)}, x_{(i)})$

$$F'(x) = \lim \frac{F(x) - F(x-1/n)}{1/n} = \frac{G_i - G_{i-1}}{x_{(i)} - x_{(i-1)}} =$$

$$\lim \frac{\Delta G}{x_{(i)} - x_{(i-1)}}$$

$$\hat{f}_i = \frac{2\Delta G}{x_{(i)} - x_{(i-1)}} - \hat{f}_{i-1} =$$

$$\frac{\Delta G}{x_{(i)} - x_{(i-1)}} + \left(\frac{\Delta G}{x_{(i)} - x_{(i-1)}} - \frac{\Delta G}{x_{(i-1)} - x_{(i-2)}} \right) -$$

$$\left(\frac{\Delta G}{x_{(i-1)} - x_{(i-2)}} - \frac{\Delta G}{x_{(i-2)} - x_{(i-3)}} \right) + \ldots \pm$$

$$\left(\frac{\Delta G}{x_{(2)} - x_{(1)}} - \frac{\Delta G}{x_{(1)} - x_{(0)}} \right)$$

Now consider the limit of the estimate. Assume:

$$\lim \left(\frac{\Delta G}{x_{(i)} - x_{(i-1)}} - \frac{\Delta G}{x_{(i-1)} - x_{(i-2)}} \right) \neq 0$$

Since the limits of both terms exist, this implies:

$$\lim \frac{\Delta G}{x_{(i)} - x_{(i-1)}} \neq \lim \frac{\Delta G}{x_{(i-1)} - x_{(i-2)}}$$

or

$$F'_+(x_{(i-1)}) \neq F'_-(x_{(i-1)})$$

Which cannot be true since $F'(x)$ exists everywhere by hypothesis. Thus

$$\lim \hat{f}_j(x) = \lim \frac{\Delta G}{x_{(i)} - x_{(i-1)}} = F'(x) \qquad \text{a.e.}$$

and by Lemma 1

$$\lim \frac{1}{d} \sum_{j=1}^{d} \hat{f}_j(x) = f(x) = F'(x) \qquad \text{a.e.}$$

The case of a repeated sample value has been lumped into the "almost everywhere" of the above proof. However, repeated sample points are easily handled by simply replacing G with qG (where q is the multiplicity of the sample point) in the formula for $\hat{f}_j(x)$. This has no effect on the proofs of the above theorems since it occurs with probability zero for finite samples from continuous distributions. In real samples we frequently encounter repeated values due to measurement inaccuracies. The above proofs still hold as long as the maximum multiplicity is less than or equal the number of subsamples. For greater multiplicity it is probably possible to show a similar result.

We now extend Theorem 2 to cover the case of certain

types of discontinuities in the density function.

Theorem 3 - The density function estimate, $\hat{f}(x)$, converges almost everywhere to the true density, $F'(x) = f(x)$, provided $F(x)$ is continuous, $f(x)$ exists, and we define $f(x) = (f(x_-)+f(x_+))/2$.

Proof - As in Theorem 2 assume we are interested in the value of $f(x)$ where $x \in [x_{(i-1)}, x_{(i)})$.

Case 1 - $f(x)$ has one discontinuity at x_0 where $x_{(i-1)} \leq x_0 \leq x_{(i)}$

$$\hat{f}_i = \frac{\Delta G}{x_{(i)}-x_{(i-1)}} + \left(\frac{\Delta G}{x_{(i)}-x_{(i-1)}} - \frac{\Delta G}{x_{(i-1)}-x_{(i-2)}}\right) -$$

$$+ \ldots \pm\left(\frac{\Delta G}{x_{(2)}-x_{(1)}} - \frac{\Delta G}{x_{(1)}-x_{(0)}}\right)$$

All the terms in parentheses above will go to zero as in Theorem 2 except those containing $(x_{(i)}-x_{(i-1)})$, leaving:

$$\lim \hat{f}_i = \lim \frac{2\Delta G}{x_{(i)}-x_{(i-1)}} - \lim \frac{\Delta G}{x_{(i-1)}-x_{(i-2)}} =$$

$$\lim \frac{2\Delta G}{x_{(i)}-x_{(i-1)}} - F'_-(x_0)$$

But if the derivative at the discontinuity is defined as the average of the left and right derivatives, then:

$$\lim \hat{f}_i = 2 \frac{F'_-(x_0) + F'_+(x_0)}{2} - F'_-(x_0) = F'_+(x_0)$$

and since no $x_{(i)} - x_{(i-1)}$ terms appear in the $\hat{f}_{i-1}$ equation, we have:

$$\lim \hat{f}_{i-1} = f(x_-)$$

That is, the value of the density estimate to the left of the discontinuity converges to the true density value to the left of the discontinuity.

Case 2 - Consider the case where the discontinuity occurs in some other interval. If it occurs after $x_{(i)}$ there is no effect and we have the same result as in Theorem 2 for $\hat{f}_i$. If the discontinuity occurs before $x_{(i)}$, say between the previous two points, we have:

$$\lim \hat{f}_i(x) = \lim \frac{\Delta G}{x_{(i)} - x_{(i-1)}} +$$

$$\lim \left(\frac{\Delta G}{x_{(i)} - x_{(i-1)}} - \frac{\Delta G}{x_{(i-1)} - x_{(i-2)}} \right) -$$

$$\lim \left(\frac{\Delta G}{x_{(i-1)} - x_{(i-2)}} - \frac{\Delta G}{x_{(i-2)} - x_{(i-3)}} \right)$$

where all other terms go to zero as in Theorem 2.

This reduces to:

$$\lim \hat{f}_1(x) = F'(x) + F'_+(x_0) - 2 \lim \frac{\Delta G}{x_{(i-1)} - x_{(i-2)}} + F'_-(x_0) =$$

$$F'(x) + F'_+(x_0) - 2 \frac{F'_-(x_0) + F'_+(x_0)}{2} + F'_-(x_0) = F'(x)$$

This proof extends directly to any finite number of discontinuities in the probability density function as long as the value at the discontinuity can be defined as the average of the values on either side. Jump discontinuities fall into this category. The only other restrictions on the estimator are:

1) The endpoint estimator must converge to the true endpoint.

2) There must be a finite number of subsamples.

Both of these restrictions are easily met.

Now that we have established the form of the estimator, we must define the following:

1) The number of subsamples
2) The choice of endpoints
3) The choice of plotting positions

Since one goal of this research is to develop a "hands-off" estimator, these choices will either be made a priori

or the estimator will make the choices based on some sample statistic.

II.3 Smoothing

Aside from the well known problems of numerical differentiation, there are two primary contributors to roughness of the density estimate as described to this point. The first is the existence of artifically large spikes due to unnaturally close spacing of several points in the random sample. The second is the tendency of the estimator itself to over (under) estimate the value of the probability density function at a point when the estimate at the previous data point was too low (high). The first problem results in a density estimate with "random" peaks and valleys, while the second tends to create oscillations in the estimate at a frequency equal to the number of sample points divided by twice the support interval. The two problems have been attacked in this dissertation somewhat independently, despite the fact that a solution to one will affect the other.

After investigating several techniques to smooth the oscillitory behavior of the estimator, including digital filtering, frequency domain modifications, and inversion of the distribution function, a straight-forward averaging technique was used. Given the data points and an estimate of the probability density function at these points,

$$\{(x_i, \hat{f}_i);\ i=0,1,\ldots,n+1\}$$

we form a new data set and corresponding set of density estimates as follows:

$$\{(y_i, \hat{f}(y_i));\ y_0=x_0;\ y_i=(x_{i-1}+x_i)/2;\ i=1,2,\ldots,n+1;$$
$$y_{n+2}=x_{n+1};\ \hat{f}(y_i)=(\hat{f}(x_{i-1})+\hat{f}(x_i))/2;\ i=1,2,\ldots,n+1;$$
$$\hat{f}(y_0)=\hat{f}_0;\ \hat{f}(y_{n+2})=\hat{f}_{n+1}\}$$

We then perform a similar procedure to get back to the original data points:

$$\{(x_i, \hat{\hat{f}}_i);\ \hat{\hat{f}}_i=(\hat{f}_i+\hat{f}_i)/2;\ i=0,1,\ldots,n;\ \hat{\hat{f}}_0=\hat{f}_0;\ \hat{\hat{f}}_{n+1}=\hat{f}_{n+1}\}$$

or after simplification:

$$\hat{\hat{f}}_i = (\hat{f}_{i-1}+2\hat{f}_i+\hat{f}_{i+1})/4 \qquad i=1,2,\ldots,n$$

By Lemma 1 this operation will not affect the convergence properties of the estimator since this is merely a convex combination of estimates which all converge to the true density.

The second type of smoothing is designed to desensitize the estimator to anomalous behavior in the data. The Quenouille-Tukey jackknife (154) and other Bootstrap methods (47,48,59) are well suited to this purpose. The fundamental technique in all of these methods is to gener-

ate estimates with portions of the data and combine them in a manner which tends to alleviate the problems associated with the estimator operating upon the entire sample. Efron (47), Rustagi (164), and Sweeder (202) have all used this approach in density estimation. The problem with these methods is that if one applies the method to a function estimate rather than a point estimate the interactions between the "subestimates" can slow the convergence of the estimator substantially.

Ideally, every random sample would be of the form:

$$x_i = F^{-1}(G_i) \qquad i=1,2,\ldots,n$$

where G_i is some plotting rule. Realistically, we are fortunate if the whole sample, let alone individual data points, accurately portrays the characteristics of the underlying density. Subsampling is a tried and proven technique to reduce the overall noisiness of the density estimate. The philosophical idea behind subsampling is to place unnaturally closely spaced data points into different subsamples before the estimate is actually developed. We have already discussed the theory; the question that remains is how many subsamples to use.

A Monte Carlo analysis was performed to determine the "optimal" subsample size. Twenty-five runs were made from each feasible combination of eight subsample sizes, (5,10, 15,20,23,25,30,45), three distributions, (uniform, normal,

and double exponential), and six sample sizes (5,10,20,40, 100,200). The mean integrated square error (MISE) and modified Cramer-von Mises (MCVM) integral error (187) were calculated and results are shown in Figure 3. As a result of this analysis the optimal points per subsample were determined to be:

Uniform type distributions.....2

Normal type distributions.....4

Laplace type distributions.....10

Actually fractionally more points per sample were used based on subsequent studies which showed that, for sample size 100, the "optimal" number of subsamples for a uniform is 46, for a normal is 23 and for a double exponential is 10. The ten subsamples for a double exponential is not really optimal in a MISE sense, but fewer subsamples were found to yield an unsatisfactorily noisy estimate while ten subsamples provided an acceptable estimate with little sacrifice in calculated error. For sample size 100 we selected 10 subsamples as the minimum number to avoid any potential noise problems.

Since the "optimal" subsample size is not a constant, we need to be able to discriminate between the classes of distributions represented by the uniform, normal and Laplace. A modification to Hogg's Q (79) statistic was chosen for this purpose.

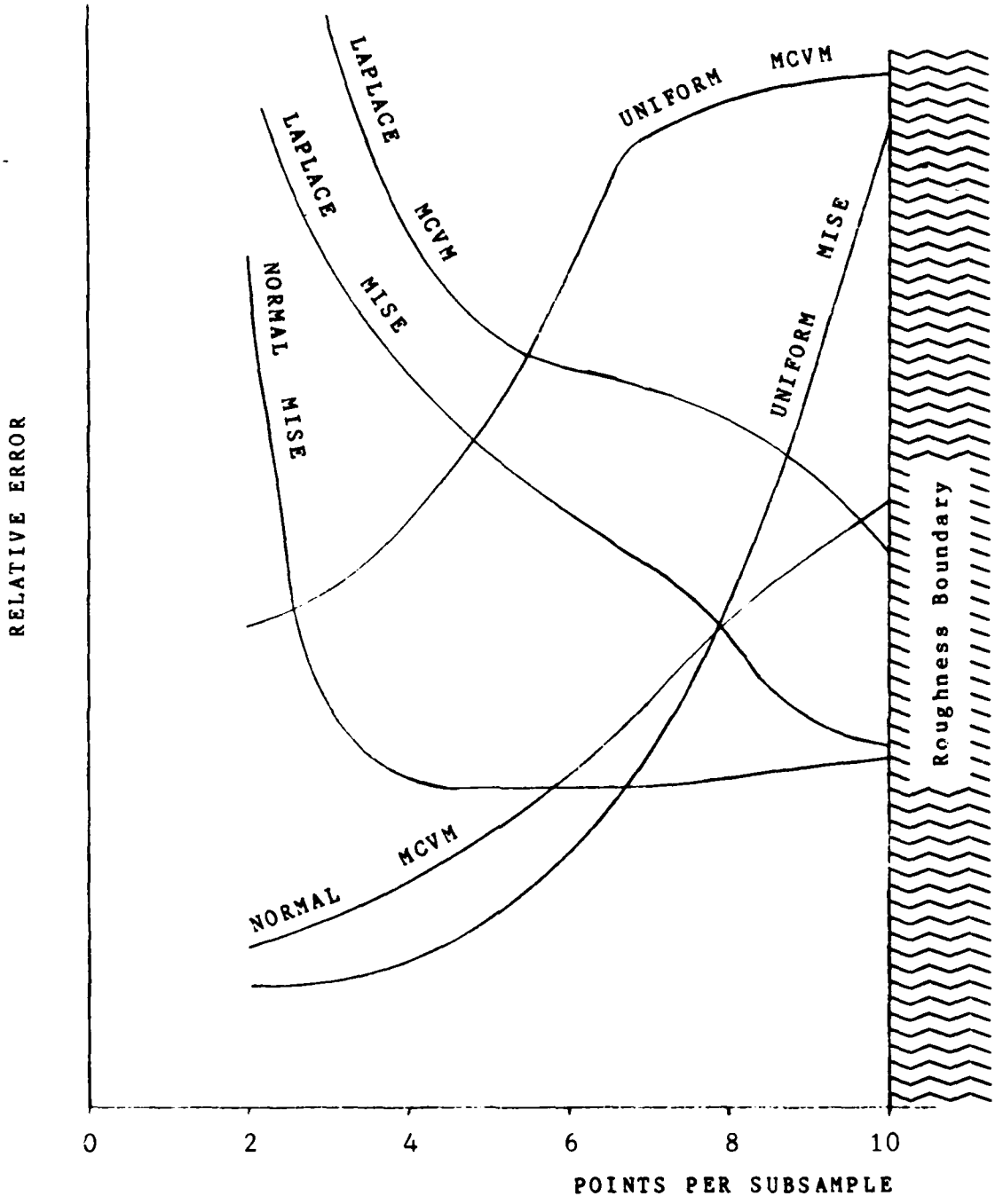


Figure 3 - Error As a Function of Subsample Size

Hogg's Q is given by:

$$Q = (U_\alpha - L_\alpha)/(U_\beta - L_\beta)$$

Where: U_α = average of largest $n\alpha$ order statistics

L_α = average of smallest $n\alpha$ order statistics

U_β, L_β are similar to U_α, L_α

The statistic defined above assumes symmetric distributions, thus it is not acceptable for a broad class of nonparametric estimators, including this one. We are particularly concerned with densities which are assymmetric or multimodal. For these purposes we define three pseudosamples:

$$\{x^1_{(i)};\ x^1_{(i)}=x_{(i)},\ x_{(i)}\leq x_m;\ x^1_{(i)}=2x_m-x_{n+1-i},\ x_{(i)}\geq x_m\}$$

$$\{x^2_{(i)};\ x^2_{(i)}=x_{(i)},\ x_{(i)}\geq x_m;\ x^2_{(i)}=2x_m-x_{n+1-i},\ x_{(i)}\leq x_m\}$$

$$\{x^3_{(i)};x^3_{(i)}=2x_{25}-x_{(n+1-i)},\ x_{(i)}\leq x_m;\ x^3_{(i)}=2x_{75}-x_{n+1-i},\ x_{(i)}\geq x_m\}$$

Where:

x_m = sample median

$x_{25} = (x_m+x_{(0)})/2$

$x_{75} = (x_m+x_{(n+1)})/2$

These pseudosamples are:

1) the first half of the original sample reflected about the median.

2) the second half of the original sample reflected about the median.

3) the first half of the sample reflected about an estimate of the 25% point and the second half reflected about an estimate of the 75% point.

The Q statistic was calculated for the original and each of the pseudo-samples (Q_0, Q_1, Q_2, Q_3). Based upon the subsample size study and the relative errors we established the following guidelines:

1) When in doubt choose too many points per subsample. An error in this case will result in the density maintaining its characteristic shape but showing noise characteristics.

2) Be absolutely certain that the density is of the uniform type before choosing the uniform, since choosing the small subsample size tends to flatten spiked densities.

In order to achieve these objectives, subsample size of about 2 was chosen only when Q_0, Q_1, Q_2, and Q_3 were smaller than the chosen breakpoint value between uniform type distributions and normal type distributions. All four values of Q were used in order to assure that the probability of a spike in any portion of the distribution was remote. Normal type distributions were assumed whenever Q_0, Q_1, and Q_2 were in the range between the uniform-normal and normal-double-exponential breakpoints. In all other cases, the distribution was assumed to be "spikey" and the subsample size, n_s, was chosen as follows:

$$n_s = \min\left(\frac{\max(Q_0,Q_1,Q_2,Q_3) - Q_n^*}{Q_d^* - Q_n^*}(n_d - n_n) + n_n \ , \ n_d\right)$$

where:

Q_n^* = theoretical Q for normal distribution

Q_d^* = theoretical Q for double exponential distribution

n_n = normal optimal subsample size

n_d = double exponential optimal subsample size

After calculating the subsample size, the calculated value was bounded on the high side by n/2 points per subsample, and on the low side by 2 points per subsample. This was based on empirical evidence that it was never advantageous to have less than two subsamples and on the inability of this estimator to calculate a density function for a sample of size one.

The values used for the breakpoints were chosen (based upon $\alpha = .04$ and $\beta = .5$) as:

$$Q_{un} = \min(1.45 + .0075n,\ 2.31)$$

$$Q_{nd} = \min(1.9 + .01n,\ 3.12)$$

which are approximate linear fits to the optimal numbers determined by Rugg (163) limited by the average population values for the distributions. These breakpoints are not critical due to the method of using pseudo-samples and based upon the relatively small variations in subsample

size. The values of population Q's used were:

$$Q_d^* = 3.53$$

$$Q_n^* = 2.70$$

$$Q_u^* = 1.92$$

This method resulted in the identifications shown in Table 1. These percentages are based on a Monte Carlo analysis of 1000 cases with sample size 100.

Table 1 - Correct Identification Percentages (n=100)

		Actual Distribution		
		Uniform	Normal	Laplace
Identified as:	Uniform	95	1	0
	Normal	5	73	0
	Intermediate	0	26	32
	Laplace	0	0	68

As the sample size decreases there is a tendency for the sample to look more like a sample from a uniform distribution. This is reflected in Table 2 which is similar to Table 1 but for sample size 10. In this case there is no intermediate subsample size due to the small number of subsamples in all cases.

Table 2 - Correct Identification Percentages (n=10)

		Actual Distribution		
		Uniform	Normal	Laplace
Identified as:	Uniform	76	34	24
	Normal	21	51	42
	Laplace	3	15	34

The amount of smoothing may be adjusted by the user if prior knowledge of the underlying density is available. However, one may easily be led into the trap of over-smoothing in order to obtain a "pretty" density while simultaneously forfeiting some accuracy.

II.4 Support Estimation

For practical purposes probability distributions can be considered to have finite support, despite the fact that they are often approximated, for mathematical convenience, by distributions with infinite support. When estimating a density function, the estimate can be quite sensitive to variation in the estimated endpoints. This is particularly true for platykurtic distributions. Consider, for example, the uniform distribution shown in Figure 4.

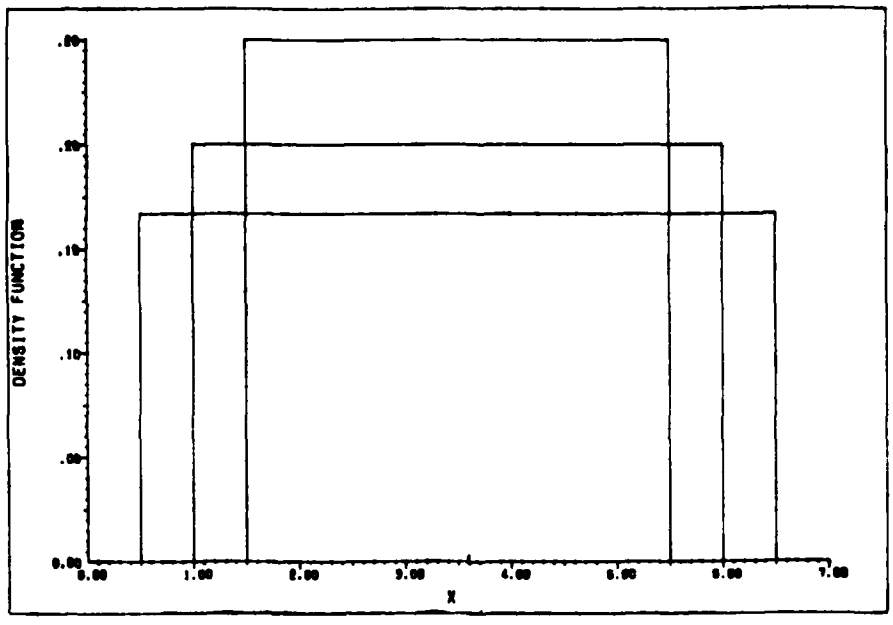


Figure 4 - Sensitivity to Support Estimation

The endpoint estimate is less critical for leptokurtic distributions (where both tails are long) since the bulk of the density function is away from the endpoint and unlikely to be greatly affected by small variations.

Endpoint selection is avoided in most non-parametric density estimation techniques by estimating $f(x|x_{(1)} \leq x \leq x_{(n)})$. Alternatives include various extrapolation rules and methods of estimating percentage points of order statistics. Hall (64) estimates the distribution of the first and last order statistic. Unfortunately, this approach is only possible with large samples, since a sample of first order statistics must be generated in order to start the estimation procedure. Bootstrap techniques (76) have been proposed for the endpoint esti-

mation task, but they frequently estimate support inside the sample bounds or well outside the actual support for samples from distributions with finite support (28,29).

The shape of the density in the vicinity of the first or last sample points is related to the distance from the extreme order statistic to the endpoint. While shape estimators of various sorts exist (kurtosis, Hogg's Q, percentile ratios) most are based on the entire sample thus somehow averaging the two tail shapes. The implicit assumption is that the distribution is symmetric. In addition, some of these statistics are quite sensitive to sample variations.

A thorough investigation was performed on a series of methods which adjust the linear extrapolation of the sample distribution function (based upon some plotting rule) to account for the estimated shape of the distribution tail. Although several of the methods developed showed a capability to predict an endpoint more accurately than a linear extrapolation, they were occasionally (less than five percent of the cases tested) drastically in error and did not, in general, perform well for small samples. The reason for lack of robustness and poor small sample performance was the paucity of information in the few sample points in the tails. The methods attempted will be described briefly as they may have some application in cases where sample sizes greater than one hundred

are available. For this estimator the new endpoint estimation techniques did not seem to significantly improve the overall performance , so a modified linear extrapolation method was used to fix the endpoints. (Only the left endpoint, $x_{(0)}$, estimate will be discussed. The right endpoint, $x_{(n+1)}$, is handled symmetrically.)

The methods investigated were:

1)

$$\hat{x}_{(0)} = 2x_{(1)} - x_{(2)}$$

Chooses as an endpoint a point the same distance to the left of the first order statistic as the second order statistic is to the right. This method has the advantage of simplicity but is extremely sensitive to sample variations. In addition, it tends to give poor results for distributions with light, long tails and for those with tails heavier than the uniform, for example a U-shaped Beta.

2)

$$\hat{x}_{(0)} = x_{(m)} - (x_{(m)}-x_{(1)})G_m/(G_m-G_1)$$

Choose as the endpoint a linear extrapolation of the points $(x_{(m)},G_m)$ and $(x_{(1)},G_1)$, $1 < m < n/2$. This method reduces the sensitivity of the estimate to sample variations but suffers from problems similar to those of method 1.

3)

$$\hat{x}_{(0)} = k_1 \sum_{i=1}^{m} (\tilde{x}-x_{(i)}) + k_2 \sum_{i=1}^{n/2} (\tilde{x}-x_{(i)})$$

$\tilde{x}$ = sample median

Chooses as the endpoint a point based on two averages relative to the sample median. This method modifies method 1 to make it more robust but still suffers from the problem of a linear estimator trying to fit a non-linear function. Method 3 also requires a relatively large sample to give reasonable results.

4)

$$\hat{x}_{(0)} = (1+kR)x_{(1)} - kR\tilde{x} \qquad 0 \leq k \leq 1$$

$$R = (n/2m) \sum_{i=1}^{m} (\tilde{x}-x_{(i)}) / \sum_{i=1}^{n/2} (\tilde{x}-x_{(i)})$$

Chooses as the estimate a linear extrapolation weighted by the function R which is a measure of the shape of the distribution similar to Hogg's Q statistic. This method is more versatile since it adjusts the slope of the extrapolation method based upon the sample. Unfortunately, the R statistic was found to be sensitive to sample variations and is not a single valued function of actual endpoint location.

5) $\hat{x}_{(0)}$ = quadratic least square fit to the points

$$(x_{(1)},G_1),\ (x_{(2)},G_2),\ (x_{(3)},G_3)$$

using the equation:

$$G(x) = a(x - \hat{x}_{(0)})^2 + b(x - \hat{x}_{(0)})$$

This method is a quadratic fit to the data subject to the constraint that the resulting equation reaches a minimum at the point $(\hat{x}_{(0)},0)$. The method is quite good for distributions with long tails but was poor for uniforms, exponentials, U-shaped betas, etc.

6)
$$\hat{x}_{(0)} = (\hat{x}_{(0)1} - K(x_{(m)}-x_{(1)}))P'$$

$$P' = \exp\{\sum_{i=1}^{m-k} [(x_{(i+k)}-x_{(i+1)})/(x_{(i+k-1)}-x_{(i)})]\} \qquad k<m$$

This method calculates a more robust percentile ratio, P', and adjusts a linear extrapolation based on the value of P'. While this method appears to have merit, in practice the value of P' was found to be non-unique for widely varying distributional shapes, and quite sensitive to sample variations due to the division and exponentiation operators.

7) Let

$$h[(\hat{x}_{(0)2}-\hat{x}_{(0)})/(x_{(m)}-x_{(1)})] = P'$$

then

$$\hat{x}_{(0)} = \hat{x}_{(0)2} - (x_{(m)}-x_{(1)})h^{-1}(P')$$

Where $\hat{x}_{(0)2} = \hat{x}_{(0)}$ as determined by method 2. This method determines the function h empirically for a series of beta distributions with various parameters and uses the inverse function to estimate the endpoint. The non-uniqueness of P′ and its sensitivity to sample variations led to poor estimates of $x_{(0)}$.

8) This method was the same as method 7 except P′ was replaced by:

$$S = (m/\ \ln(n)) \sum_{i=1}^{m} (x_{(m+1)}-x_{(i)})^4 / [\sum_{i=1}^{m} (x_{(m+1)}-x_{(i)})^2]^2$$

The method is inspired by the sample kurtosis with an empirically defined scaling factor, m/ ln(n), included to reduce sensitivity to sample size and the fractional portion of the sample, m/n, used in the calculation of S.

Method 1 was used in this estimator (modified as described below) for the following reasons:

1) The ability to generate a density estimate for small samples was desired. All other endpoint estimation schemes require larger samples than method 1 to give reasonable results.

2) The method is simple with no subtle pitfalls and gives reasonable results which do not contradict known facts.

Philosophically, one feels that the entire sample, rather than a subsample, must be "better" for approximating endpoints. The problem with picking endpoints and then using the same selected values in the calculations of each of the subsample densities is that too much probability tends to be lumped in the tails. On the other hand, allowing each subsample to determine its own endpoints tends to spread the estimate over a wider support which adversely affects estimator performance for densities which do not tend to zero as the endpoint is approached. A compromise solution was developed experimentally which seems to eliminate both these problems.

Define:

$\hat{x}_{(0)}$, $\hat{x}_{(n+1)}$ = estimates of endpoints based upon the entire sample

$\hat{x}_{(0)i}$, $\hat{x}_{(n+1)i}$ = estimates of endpoints based upon the i^{th} subsample

$\hat{x}^{*}_{(0)i}$, $\hat{x}^{*}_{(n+1)i}$ = endpoint estimate used in calculating density from the i^{th} subsample

Then

$$\hat{x}^{*}_{(0)i} = \max(\hat{x}_{(0)i}, \hat{x}_{(0)})$$

$$\hat{x}^{*}_{(n+1)i} = \min(\hat{x}_{(n+1)i}, \hat{x}_{(n+1)})$$

As can be seen in Chapter IV.2, this endpoint estimation technique, when applied to small samples, results in

an excellent approximation to the 1/n and (n-1)/n percentage points of the true distribution by the estimated distribution. Thus the errors in the endpoint estimates due to this method are insignificant for most applications other than approximating points far out in the tails of the distribution.

II.5 Plotting Position Selection

Plotting positions are defined as a set of cumulative probabilities associated with a set of ordered observations. Their purpose stems from the use of probability paper (as far back as 1896) to try to predict distributions of observed random variables. They were commonly used by hydrologists to analyze flood data (74). Generally an attempt is made to approximate some point in a distribution by choice of plotting position, for instance $E[F(x_i)]$.

As Harter (69) points out in his excellent summary of the history and use of plotting positions, much of the problem regarding the choice of plotting positions is due to the fact that

$$F[E(x_i)] \neq E[F(x_i)] = i/(n+1)$$

except for a uniform distribution. The median ranks choice of plotting position is attractive for the case of a single point, since for monotonic functions the median

of the function is the function of the median. Unfortunately this is not true for functions of more than one random variable.

Table 3 shows some of the historically more popular choices of plotting positions. A small amount of empirical investigation of many of these plotting positions was done but there was no obviously better choice for the determination of the density estimate. Harter (69) gives

Table 3 - Plotting positions of the i^{th} Order Statistic

F(x)	Description
1. i/n	value of the empirical distribution function (EDF)
2. $i/(n+1)$	mean rank
3. $(i-1)/(n+1)$	mode rank
4. $(i-.3)/(n+.4)$	median rank (approximation)
5. $(i-.5)/n$	midpoint of the jump of the EDF
6. $(n(2i-1)-1)/(n-1)$	average of mean and mode ranks
7. $(i-.375)/(n+.25)$	efficient approximation for the normal distribution
8. $(i-a)/(n-a-b+1)$ $a,b \leq 1$	Blom's plotting position (15)
9. $(i+a)/(n+b)$ $-1 \leq a \leq b \leq 1$	a more general plotting position

a detailed analysis of choices of plotting rules and that will not be repeated here. For the purposes of this work, the approximate median ranks:

$$G_i = (i-.3)/(n+.4)$$

plotting positions were used. This is the same approach taken by Sweeder.

While the plotting rule does not in itself greatly affect the estimate, in conjunction with subsampling it does. The reason for this is that equal areas are forced into unequal intervals at the ends of the subsample. Figures 6 and 7 illustrate this problem. In Figure 5 we have a sample shown along with the density generated with no subsampling. If we subsample twice we obtain Figures 6 and 7 which are averaged to get the smoothed estimate, Figure 8. Note that the subsamples and the resulting estimate tend to have peaks near the endpoints. This is due to forcing the estimate to generate too much area at the ends. For example, in subsample one, approximately one-fifth the area is generated between $x_{(0)}$ and $x_{(1)}$, while this interval is only one of the nine defined by the endpoints and sample points. Sweeder's estimates frequently showed a characteristic hump near the endpoints, which was due to this phenomenon.

A solution to this problem is to reapportion the area generated between the points of the subsample. A new set

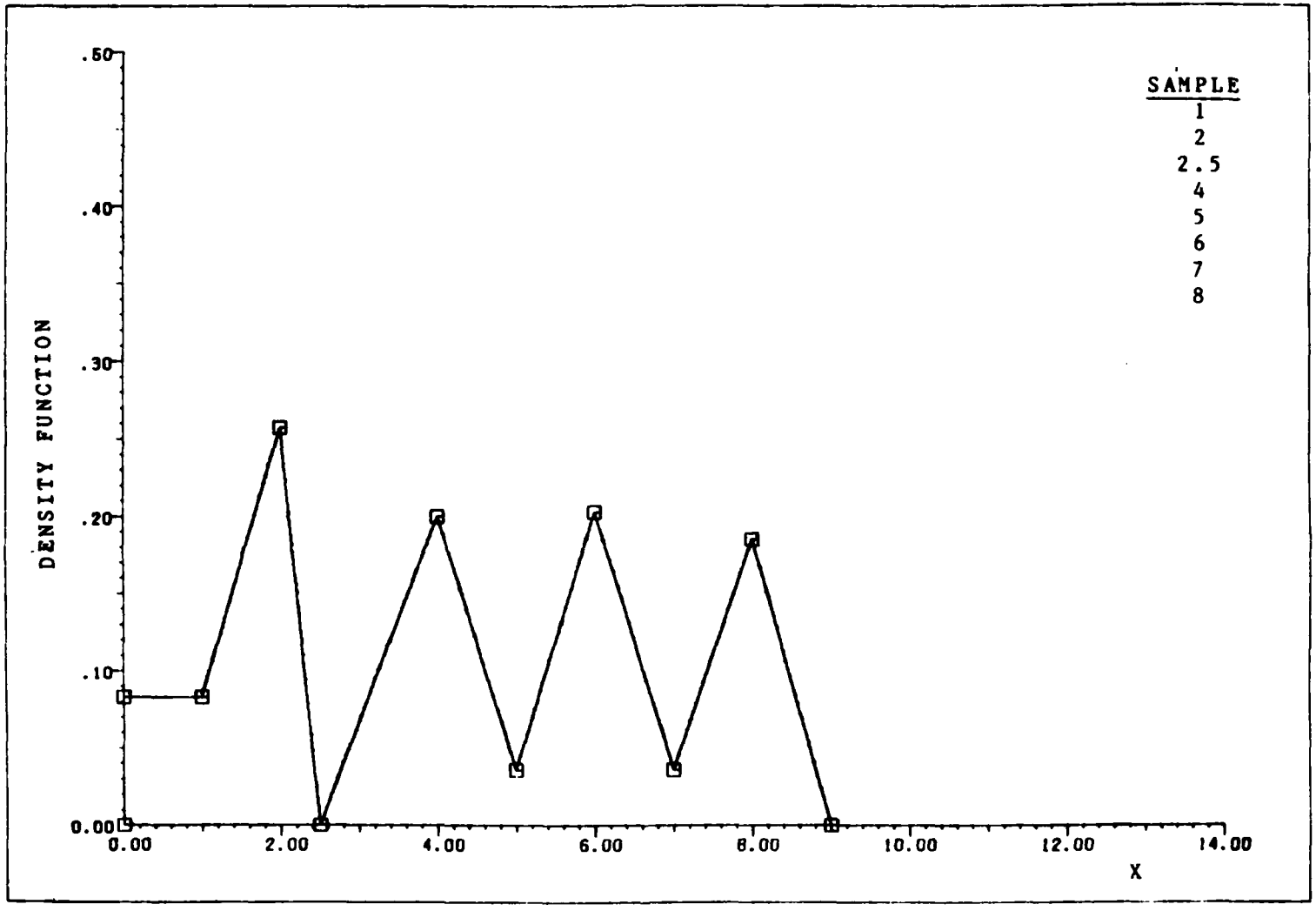


Figure 5 - Estimate of Density Function With No Subsampling

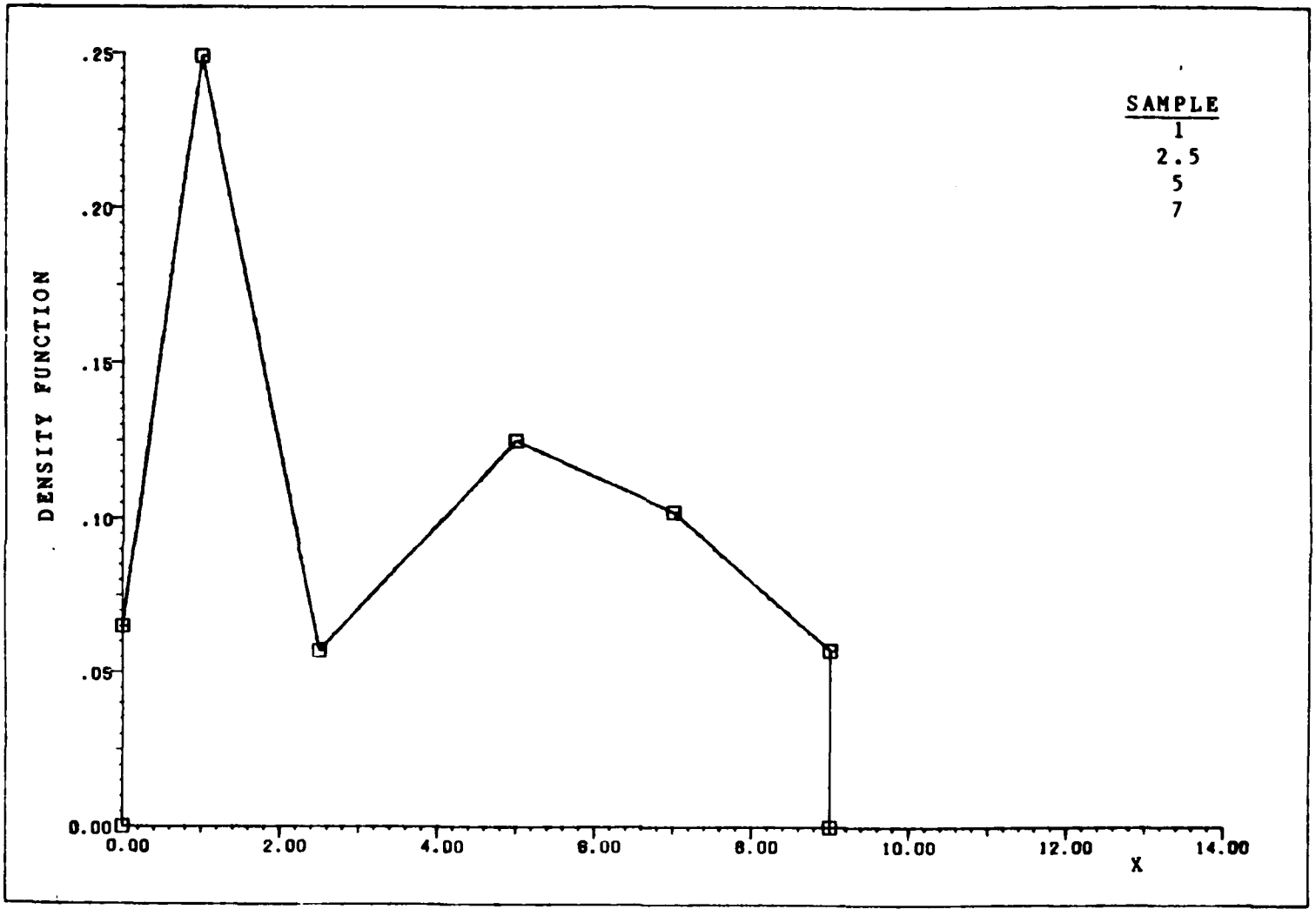


Figure 6 - Density Estimate Generated from Subsample One

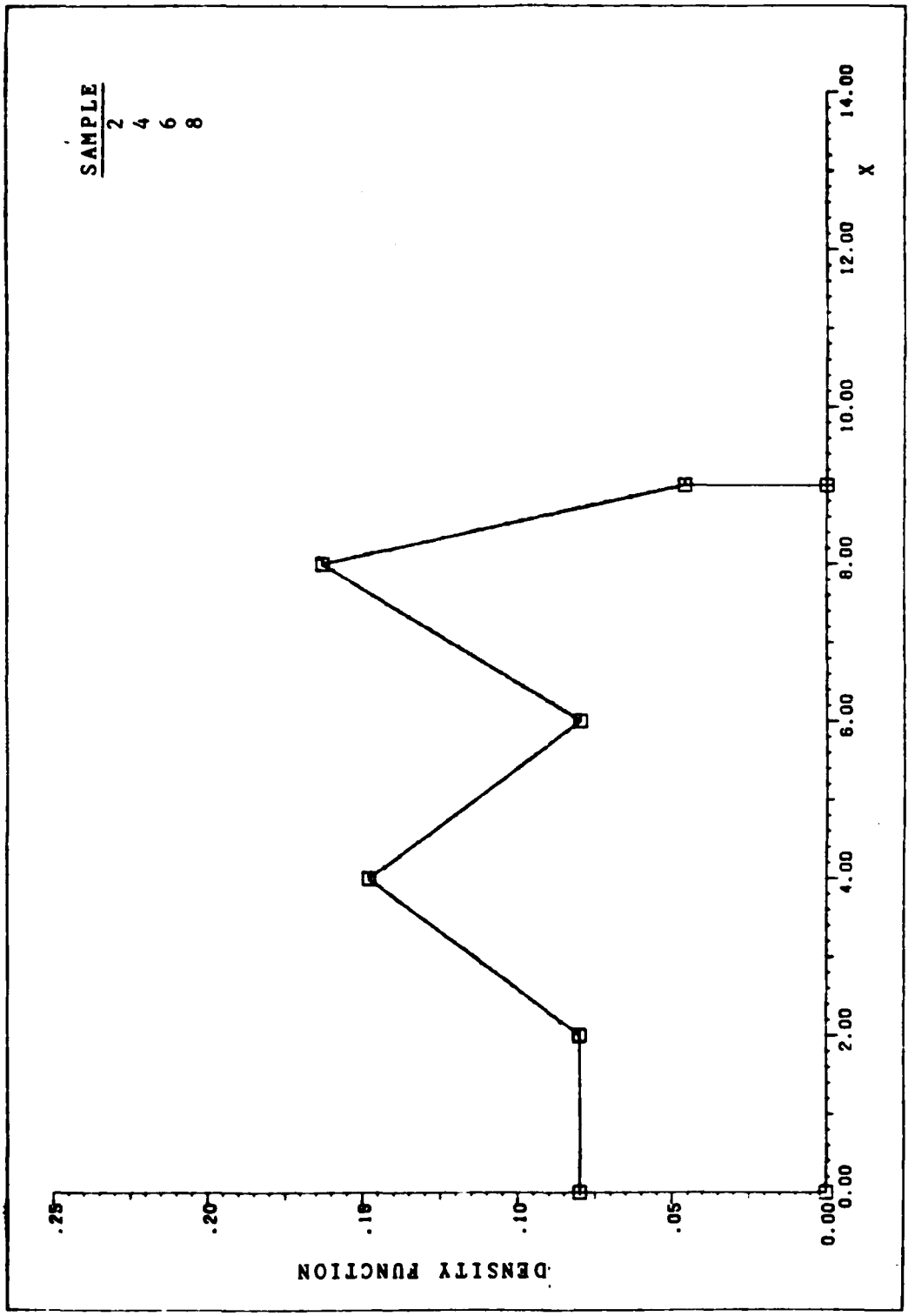


Figure 7 - Density Estimate Generated from Subsample Two

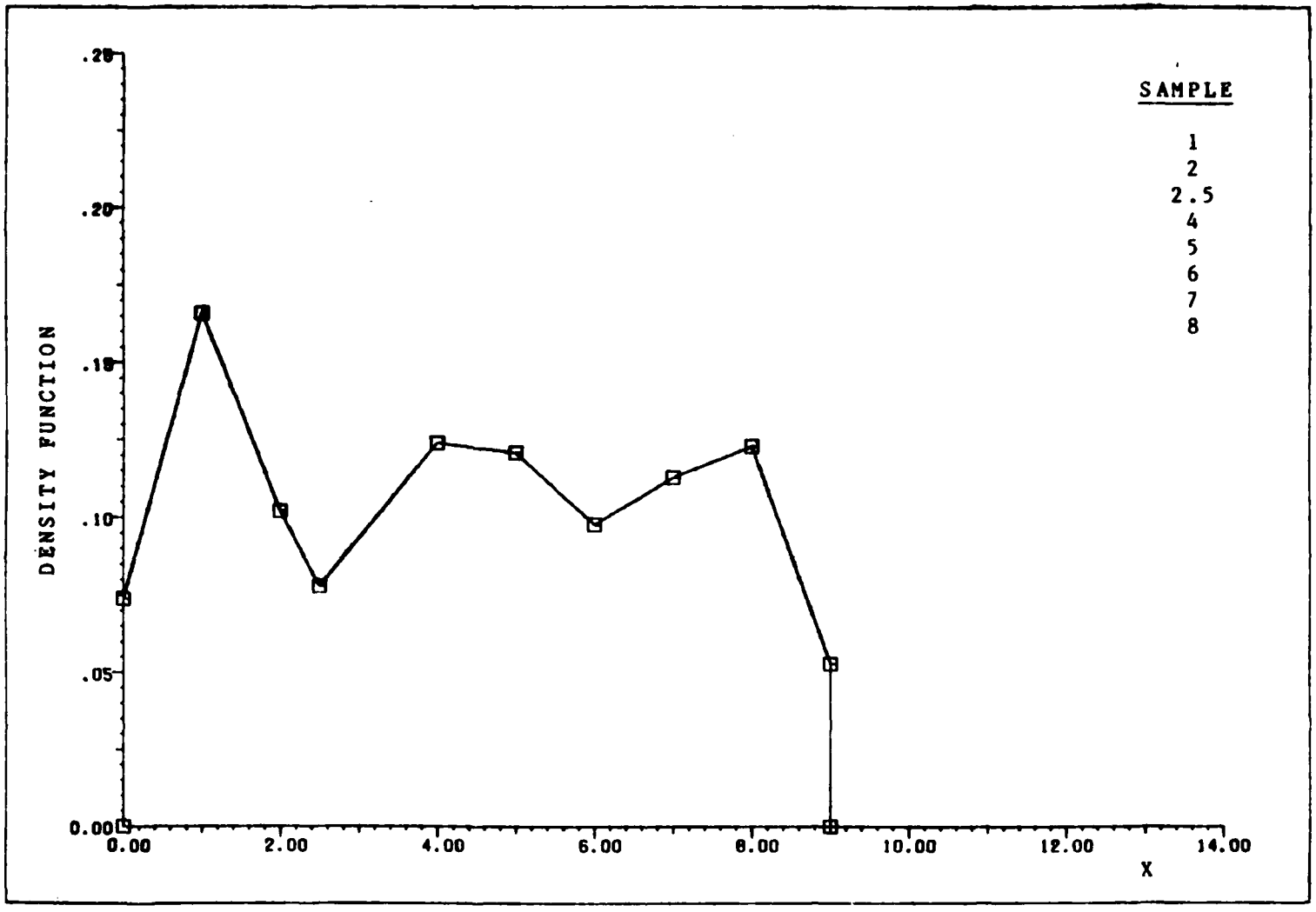


Figure 8 - Smoothed (by Subsampling) Density Estimate

of plotting points are chosen such that the values of G_{1s} and G_{ns}, the first and last subsample plotting positions, are equal to the plotting positions for the corresponding point in the entire sample. The plotting positions for the rest of the subsample points are determined by simply dividing $G_{1s}-G_{ns}$ by the number of intervals remaining. This has the result of making the estimated subsample probability density function values more closely represent the entire sample density function in the tails, while taking advantage of the smoothing properties of subsampling throughout the rest of the support. Since this approach is equivalent to selecting a different set of (assymetric) plotting positions, the convergence properties of the estimator remain unchanged.

II.6 Example Problem

The following example illustrates the use of the new density estimator. Data used represents the lifetimes of eight grinding wheels and are extracted from Table 11.10 of Kapur and Lamberson (88).

$$X = (22,25,30,33,35,52,63,104)$$

First we will calculate an estimate with no subsampling or smoothing to illustrate the technique. Following this we will calculate the smoothed estimate as described in the earlier portions of this Chapter.

The plotting positions are:

$$G = \frac{i-.3}{n+.4} = (.0833, .2024, .3214, .4405,$$

$$.5595, .6786, .7976, .9167)$$

Endpoints are calculated as:

$$x_0 = 19 \qquad x_9 = 145$$

The forward pass:

$$f_0 = 0$$

$$f_1 \Delta x_1/2 = .0833 \Rightarrow f_1 = .0238$$

$$f_2 = .0238$$

$$f_3 = .0238$$

$$f_4 = .0555$$

$$f_5 = .0635$$

$$f_6 < 0$$

So we set $f_6 = 0$

And recalculate $f_5 = (.476 - f_4(x_4-x_3))/(x_5-x_3) = .0163$

Continuing $f_7 = .0216$

$$f_8 < 0$$

So we set $f_8 = 0$

And recalculate $f_7 = .0092$

$$f_9 = .0041$$

We now calculate the backward pass:

$$f_9 = 0$$

$$f_8 \Delta x_9/2 = .0833 \Rightarrow f_8 = .0041$$

$$f_7 = .0017$$

$$f_6 = .0199$$

$$f_5 < 0 \Rightarrow f_5 = 0$$

Recalculate

$$f_6 = .0163$$

$$f_4 = .119$$

$$f_3 < 0 \Rightarrow f_3 = 0$$

Recalculate

$$f_4 = .0952$$

$$f_2 = .0476$$

$$f_1 = .0317$$

$$f_0 = .0238$$

Averaging the forward and backward passes yields:

$$\hat{f} = (.0119,.0436,.0357,.0119,.0754,.0082,$$
$$.0082,.0055,.0021,.0021)$$

The result of this estimation is shown in Figure 9. Note the amount of noise in even this simple estimate.

We now calculate the "optimal" number of subsamples as two and obtain the smoothed estimate.

$$Y_1 = (22,30,35,63)$$

$$Y_2 = (25,33,52,104)$$

$$y_{1_0} = \max(19,14) = 19 \qquad y_{1_5} = \min(91,145) = 91$$

$$y_{2_0} = \max(19,17) = 19 \qquad y_{2_5} = \min(145,156) = 145$$

The first subsample is augmented with the endpoints

$$(19,22,30,35,63,91)$$

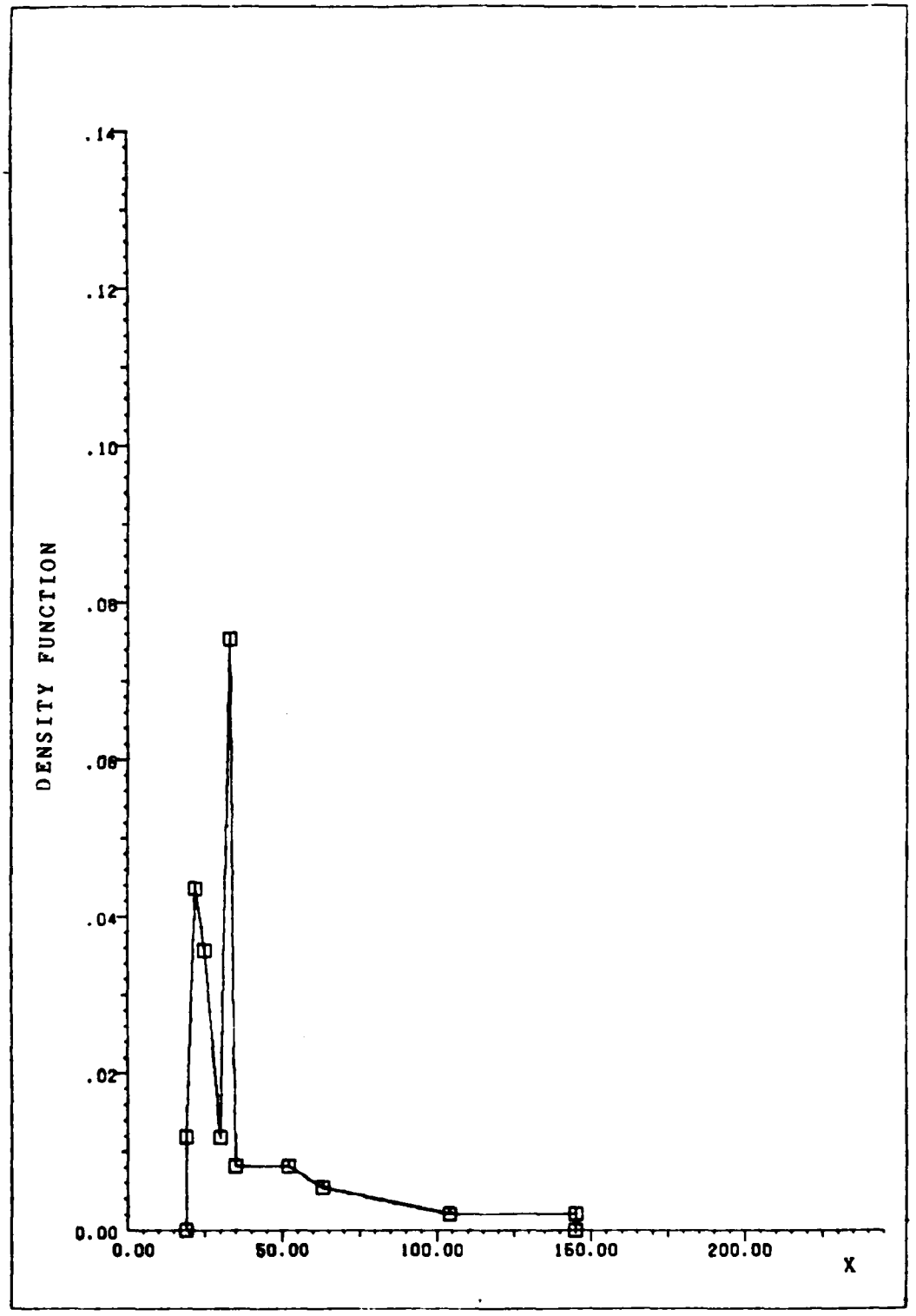


Figure 9 - Example Density Estimate Before Smoothing

Plotting positions are calculated

$$G_1 = .7/8.4 = .0833 \qquad G_4 = 3.7/4.4 = .8409$$

$$G_2 = (G_4+2G_1)/3 = .3358 \qquad G_3 = (2G_4+G_1)/3 = .5884$$

Forward pass is calculated as before:

$$f_0 = 0$$
$$f_1 = .0555$$
$$f_2 = .0076$$
$$f_3 = .0934$$
$$f_4 < 0 \Rightarrow f_4 = 0$$

Recalculate

$$f_3 = .0292$$
$$f_5 = .0114$$

Backward pass is now calculated:

$$f_5 = 0$$
$$f_4 = .0114$$
$$f_3 = .0066$$
$$f_2 = .0944$$
$$f_1 < 0 \Rightarrow f_1 = 0$$

Recalculate

$$f_2 = .0763$$
$$f_0 = .0555$$

The first subsample estimate is given by the pairs:

(19,.0278) (22,.0278) (30,.0420)

(35,.0179) (63,.0057) (91,.0057)

Performing similar calculations for the second subsample we obtain the pairs:

(19,.0055) (25,.0476) (33,.0156)
(52,.0086) (104,.0021) (145,.0021)

We now smooth the estimates using

$$\hat{\hat{f}}_i = (\hat{f}_{i-1} + 2\hat{f}_i + \hat{f}_{i+1})/4$$

to obtain, for the first subsample:

(19,.0278) (22,.0314) (30,.0324)
(35,.0209) (63,.0088) (91,.0057)

and for the second subsample:

(19,.0055) (25,.0291) (33,.0219)
(52,.0087) (104,.0037) (145,.0021)

Note that at this point the density estimates no longer integrate to one since the smoothing operation is not weighted by the sample intervals. This will be corrected after averaging, but first we must interpolate within each sample to find the values at corresponding x coordinates.

The first subsample provides:

(19,.0278) (22,.0314) (25,.0318) (30,.0324)
(33,.0255) (35,.0209) (52,.0136) (63,.0088)
(91,.0057) (104,0.0) (145,0.0)

Note that there is additional area added implicitly between the points with x coordinates of 91 and 104. This provides additional smoothing for the transition between subsample estimates.

The second subsample provides:

(19,.0055)	(22,.0173)	(25,.0291)	(30,.0240)
(33,.0219)	(35,.0205)	(52,.0087)	(63,.0076)
(91,.0049)	(104,.0037)	(145,.0021)	

Averaging the two subsample estimates we obtain:

(19,.0167)	(22,.0244)	(25,.0305)	(30,.0282)
(33,.0237)	(35,.0207)	(52,.0112)	(63,.0082)
(91,.0053)	(104,.0019)	(145,.0011)	

Integrating, we see that the total area under the curve is 1.08815 so we divide each of the pdf estimates by this value to obtain our final estimate:

(19,.0153)	(22,.0224)	(25,.0280)	(30,.0259)
(33,.0218)	(35,.0190)	(52,.0103)	(63,.0075)
(91,.0049)	(104,.0017)	(145,.0010)	

Figure 10 shows this estimate. Also shown is the Weibull fit, determined using Weibull probability paper, given by Kapur and Lamberson. Since the two density estimates are clearly different, the natural question is which is better. There is no answer to this question, however we

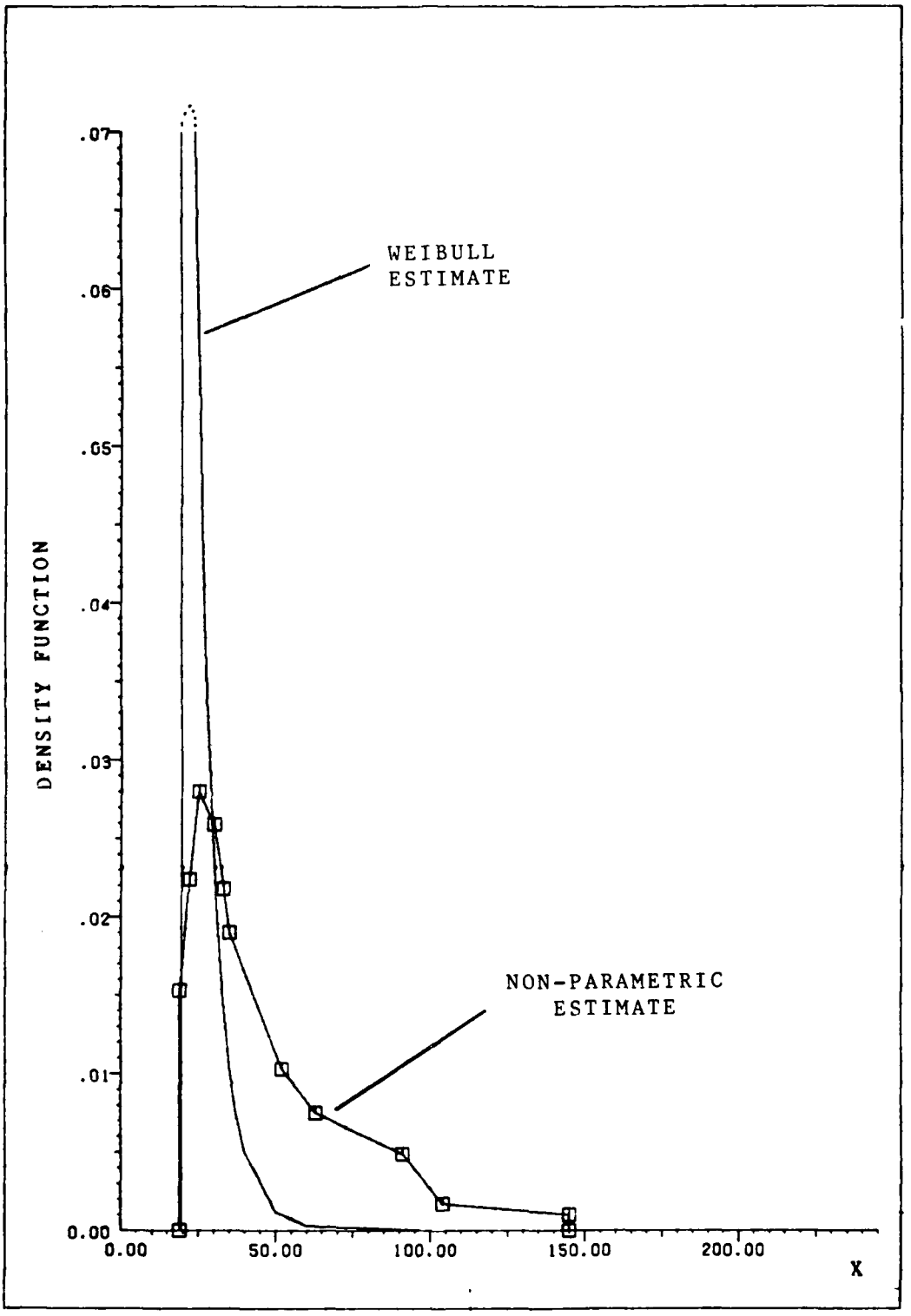


Figure 10 - Example Density Estimate

can calculate the likelihood of the sample for each of the distributions (assuming independent sample points.) When we do this the value for the Weibull estimate, L_W, and the value for the new estimate, L, are:

$$L_W = 9.5 \times 10^{-21}$$

$$L = 8.4 \times 10^{-16}$$

Clearly this particular sample is much more likely to be from the new density estimate. It is also interesting to calculate the likelihood of this sample coming from a uniform distribution, since, for very small samples, a *uniform distribution* tends to maximize the likelihood. The likelihood for a uniform distribution is:

$$L_U < 4.9 \times 10^{-16}$$

One other question remains along these lines and that is the question of what the subsampling and smoothing have done to the likelihood of the sample. If we calculate the likelihood of this sample for the unsmoothed estimate with no subsampling we obtain:

$$L = 1.1 \times 10^{-15}$$

The significance of these numbers is subjective. They are presented more as a matter of interest than as a claim for the quality of the new estimator.

III. Quality of the Estimator.

Now that we have developed the new non-parametric estimator, an evaluation of the quality of this estimator is required. The approach taken in this chapter is to obtain Monte Carlo estimates of the accuracy of the new estimator and compare these results with existing density estimators. Wegman (226) provides a discussion of and results from several other non-parametric density estimates. Tapia and Thompson (203) discuss properties of some other estimators and Sweeder (202) also provides some comparisons. In order to compare with these other results, twenty five Monte Carlo repetitions using samples of size one hundred were used in this study. The results presented here are consistent with Monte Carlo studies using one hundred repetitions.

The measure of merit most frequently used in these comparisons is Mean Integrated Square Error (MISE) or frequently MISE is approximated by the average squared error (ASE). The following measures of merit were considered as alternatives in this study:

1) Anderson-Darling Integral
2) Modified Anderson-Darling Integral
3) Average Square Error
4) Cramer-von Mises Integral
5) Modified Cramer-von Mises Integral

6) Kolmogorov-Smirnov Integral

7) Modified Kolmogorov-Smirnov Integral

8) Integrated Square Error

9) Integrated Absolute Error

10) Maximum Absolute Deviation

Although all of the above measures gave different numbers, relative comparisons indicated that only Maximum Absolute Deviation varied significantly from the others. This was true primarily at the points of discontinuity in the density function. Thus, for the purposes of this study, Average Square Error (approximately equal to MISE) is used when comparisons are made with other density estimators, and Integrated Absolute Error is used for two sample tests since the numbers retain more of an intuitive feel, at least to the author. Based upon results of investigations of those ten measures of merit, the research results would not change substantially if any of the first nine measures were used and possibly not even for the last. For derivations and definitions of these error measures see Sweeder or Sahler (165).

Both the estimated probability density function and cumulative distribution function are compared with other estimation techniques. Tables 4 and 5 show the results of these comparisons. The results of the density function comparisons indicate that the new estimator is clearly superior for platykurtic distributions, equal to the best

	H* Histogram I	H* Histogram II	K* Kernel	T* Trigonometric	S+ Sweeder 4	S+ Sweeder 5	S+ Sweeder 6	New Estimator
Uniform (0,1)	.1210 (.0960)	.1544 (.1162)	.0439 (.0187)	.0297 (.0480)	.0644 (.0199)	.0627 (.0188)	.0596 (.0246)	.0048 (.0031)
Normal	.0054 (.0034)	.0146 (.0127)	.0012 (.0010)	.0012 (.0012)	.0021 (.0018)	.0023 (.0019)	.0018 (.0016)	.0012 (.0006)
Laplace	-	-	-	-	.0025 (.0015)	.0025 (.0015)	.0025 (.0015)	.0059 (.0035)

Standard Error in Parentheses

* Wegman(226)

+ Sweeder (202)

Table 4 - Comparison of Density Estimators, Average Square Error (n=100)

	Empirical D. F.	Sweeder 4+	Sweeder 5+	Sweeder 6+	New Estimator
Uniform (0,1)	.00167	.00105 (.00074)	.00106 (.00080)	.00093 (.00076)	.00104 (.00070)
Normal	.00167	.00131 (.00139)	.00136 (.00139)	.00126 (.00138)	.00054 (.00015)
Laplace	.00167	.00080 (.00078)	.00080 (.00078)	.00085 (.00080)	.00151 (.00049)

Standard Error in Parentheses

+ Sweeder (202)

Table 5 - Comparison of Distribution Estimators, Average Square Error (n=100)

previous estimates for mesokurtic distributions, and slightly inferior to Sweeders for leptokurtic distributions. A small number of cases have been run in each of the distribution classes with Cauchy, exponential, and various beta distributions. These additional runs show that uniform, normal, and double exponential distributions are indeed representative of these other distributions.

The plots shown in Figures 11 to 16 demonstrate graphically the ability of the estimator to fit various densities. The normal, uniform, and double exponential plots show the maximum likelihood estimate (parametric) for the random sample. Actual data show that the nonparametric estimate has smaller MISE than the parametric estimate for the normal 31% of the time for sample size 100. These plots are the best results obtained from one hundred random samples from each distribution. The following plots, Figures 17 to 21, are the best results obtained from ten samples, each of size one hundred, from the distributions shown.

The results for the distribution function are quite similar, showing consistent improvement over the empirical distribution function, and errors comparable to the extremely good values achieved by Sweeder. Note that the estimator was "optimized" to provide a good density estimate. Improvement in the estimate of the distribution

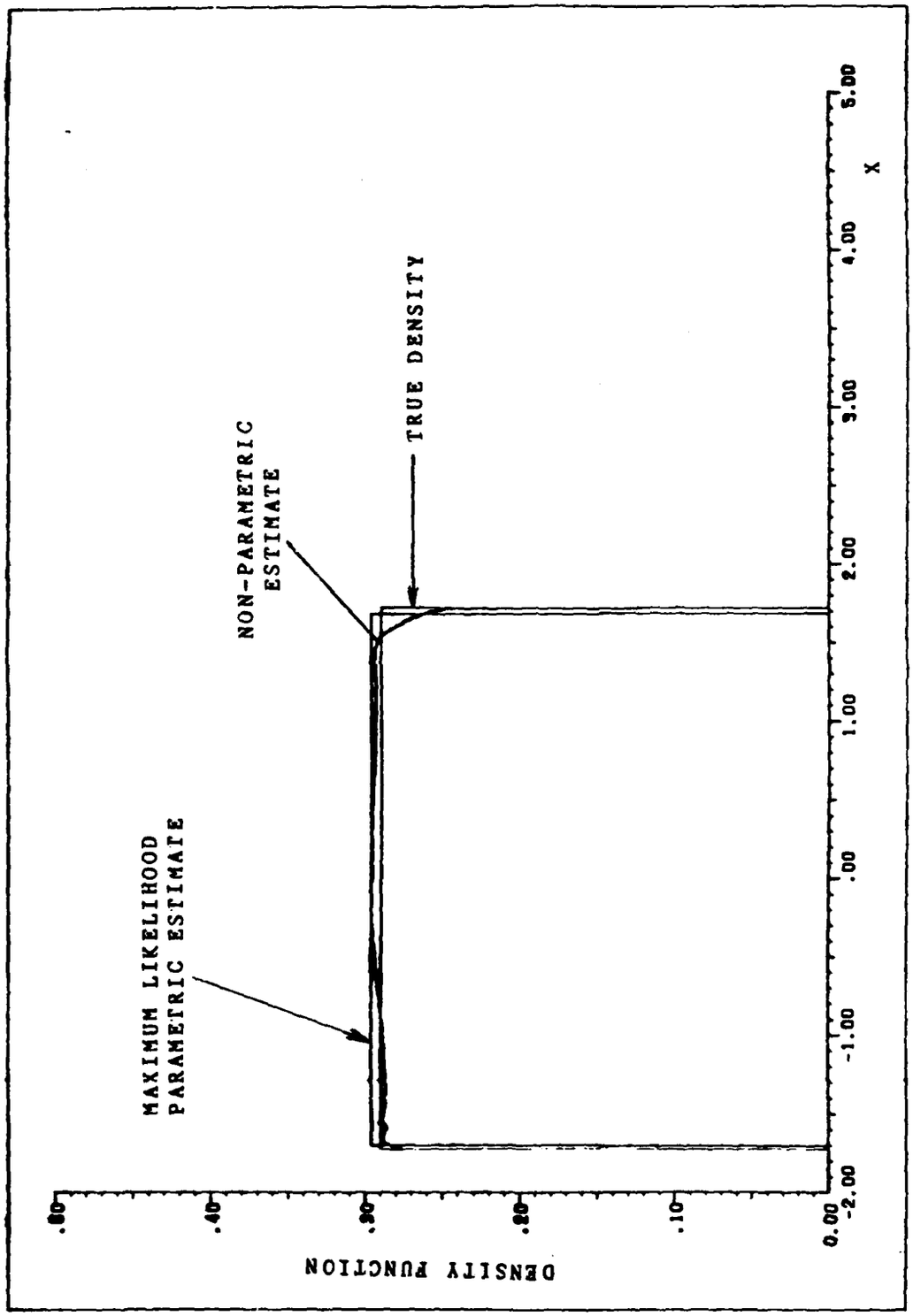


Figure 11 - Uniform Density Estimate (n=100)

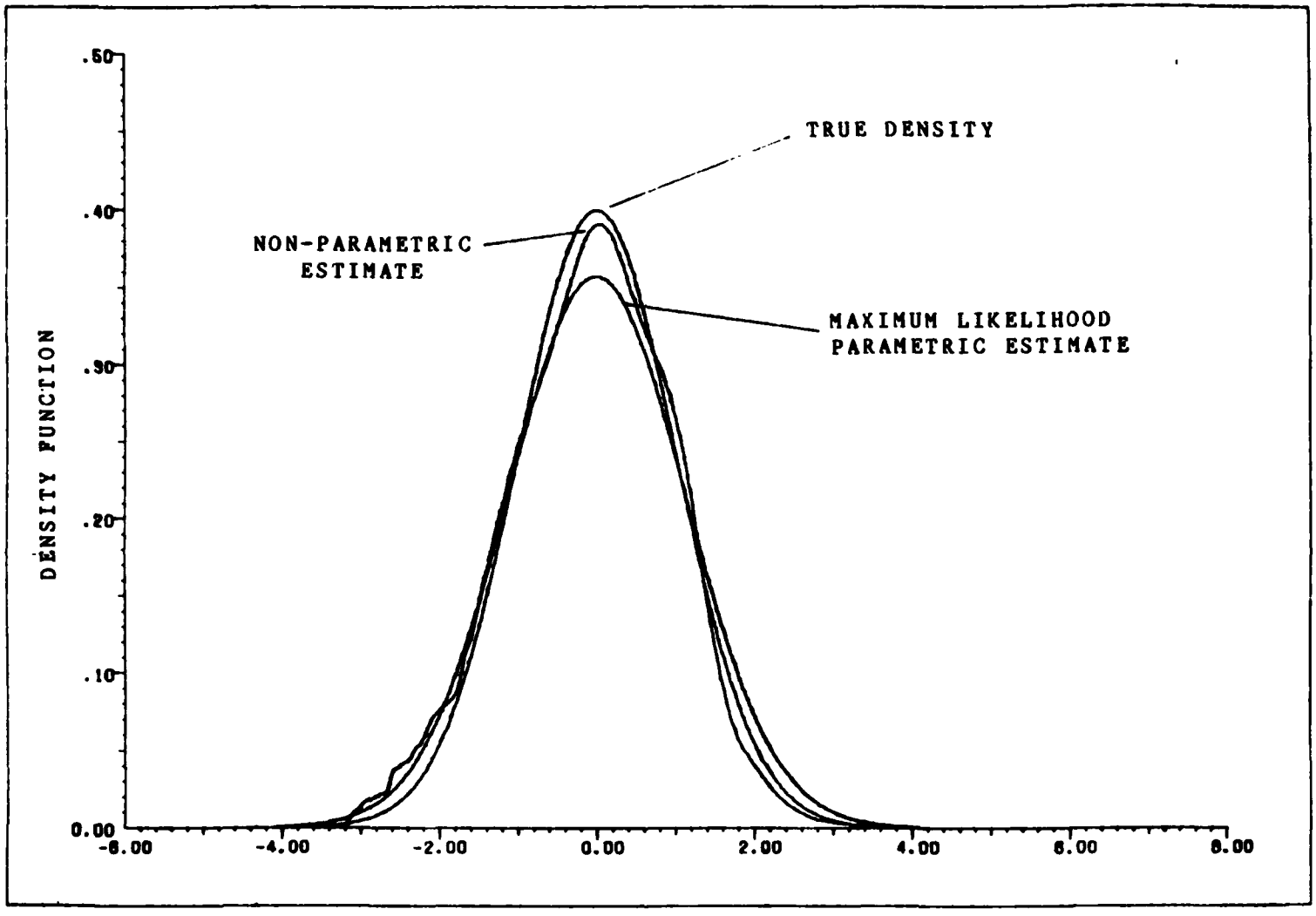


Figure 12 - Normal Density Estimate (n=100)

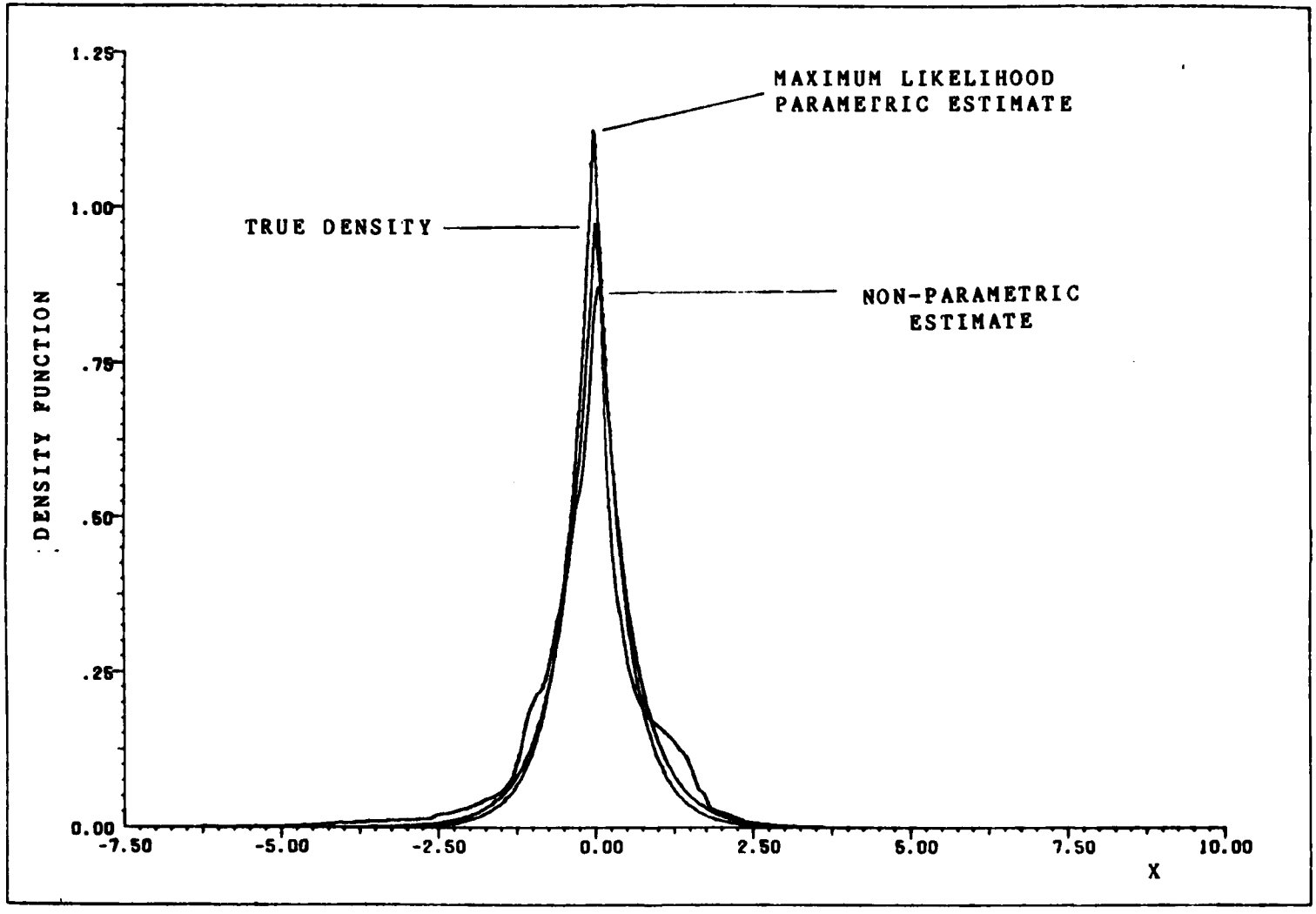


Figure 13 - Laplace Density Estimate (n=100)

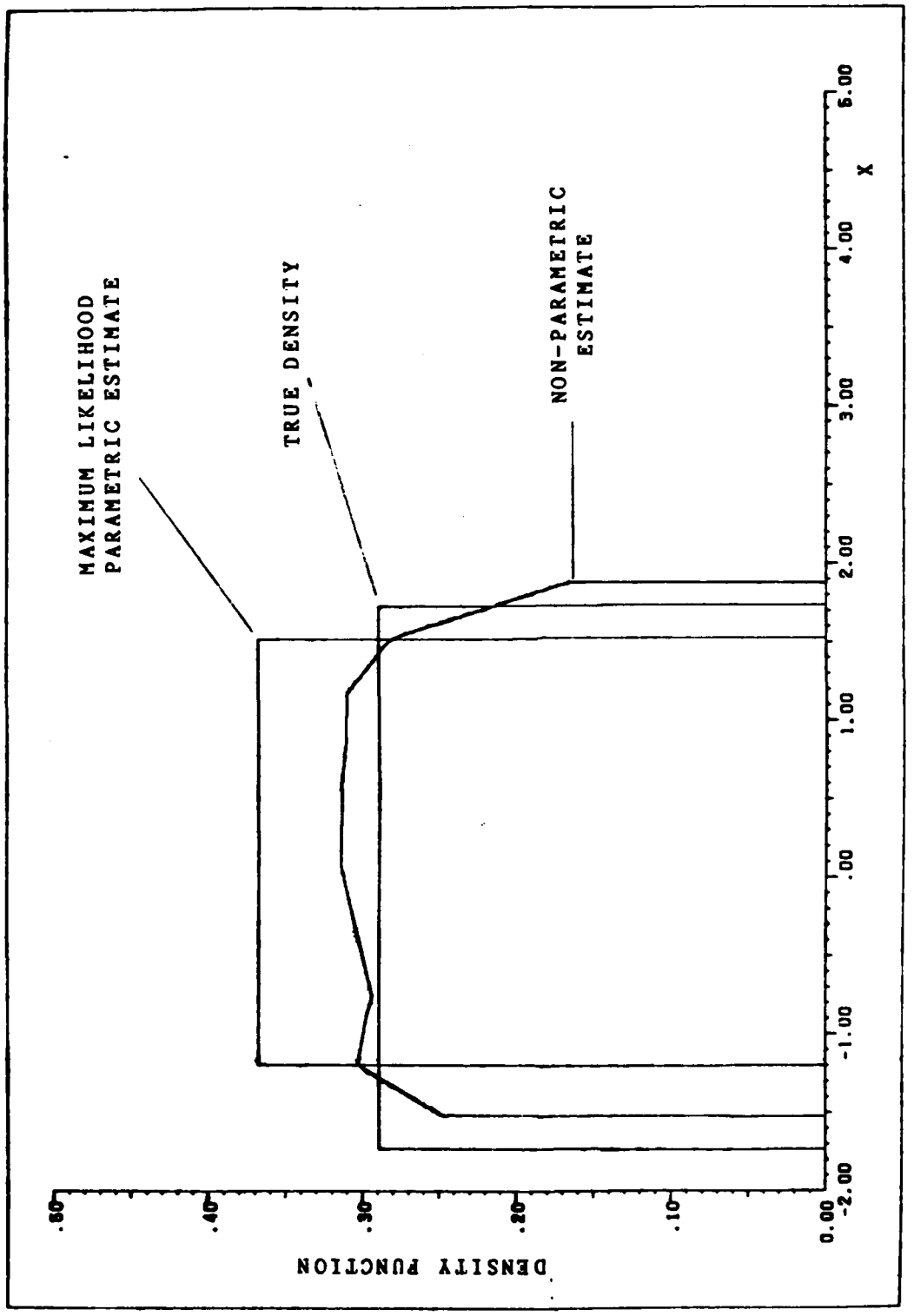


Figure 14 - Uniform Density Estimate (n=10)

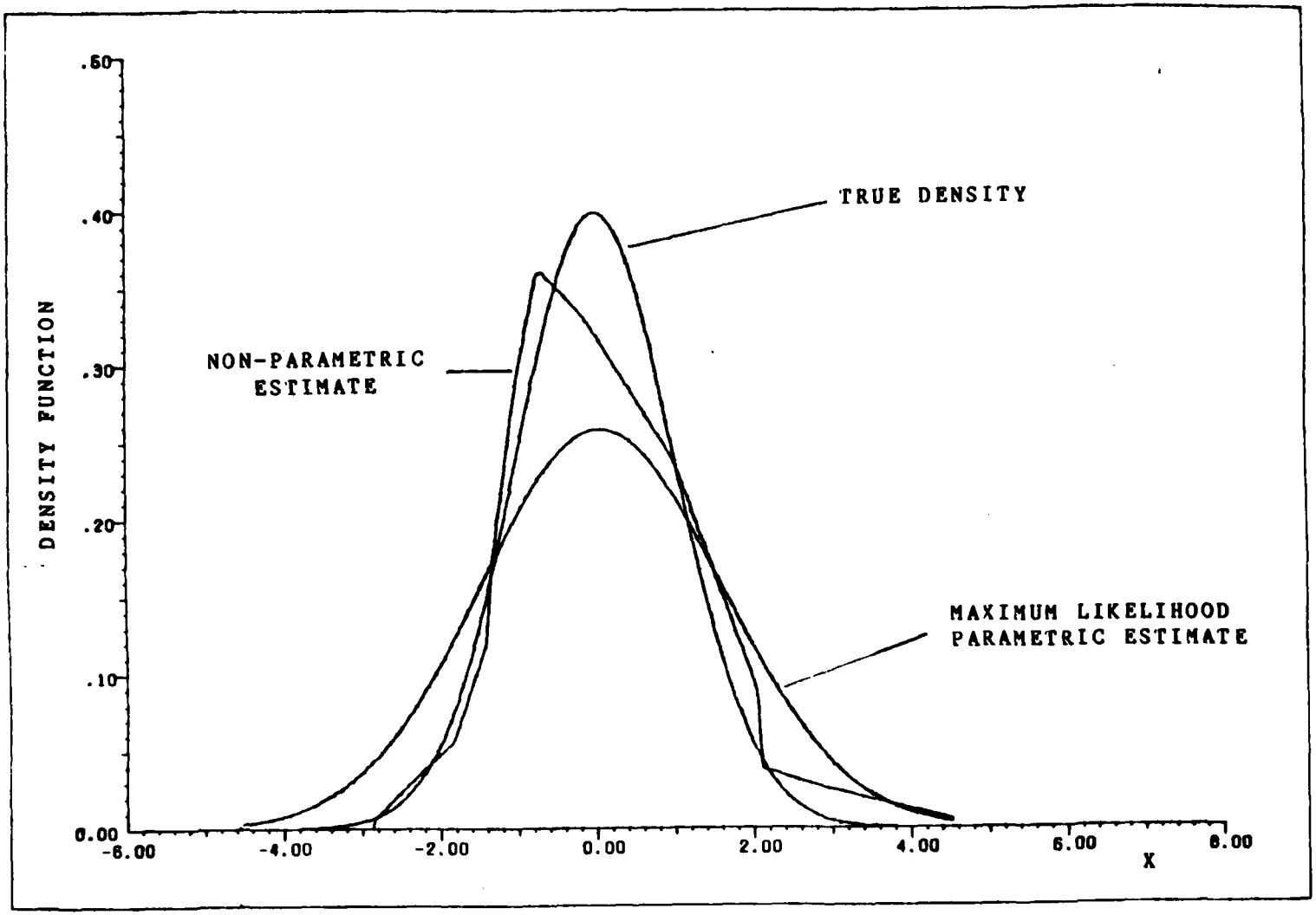


Figure 15 - Normal Density Estimate (n=10)

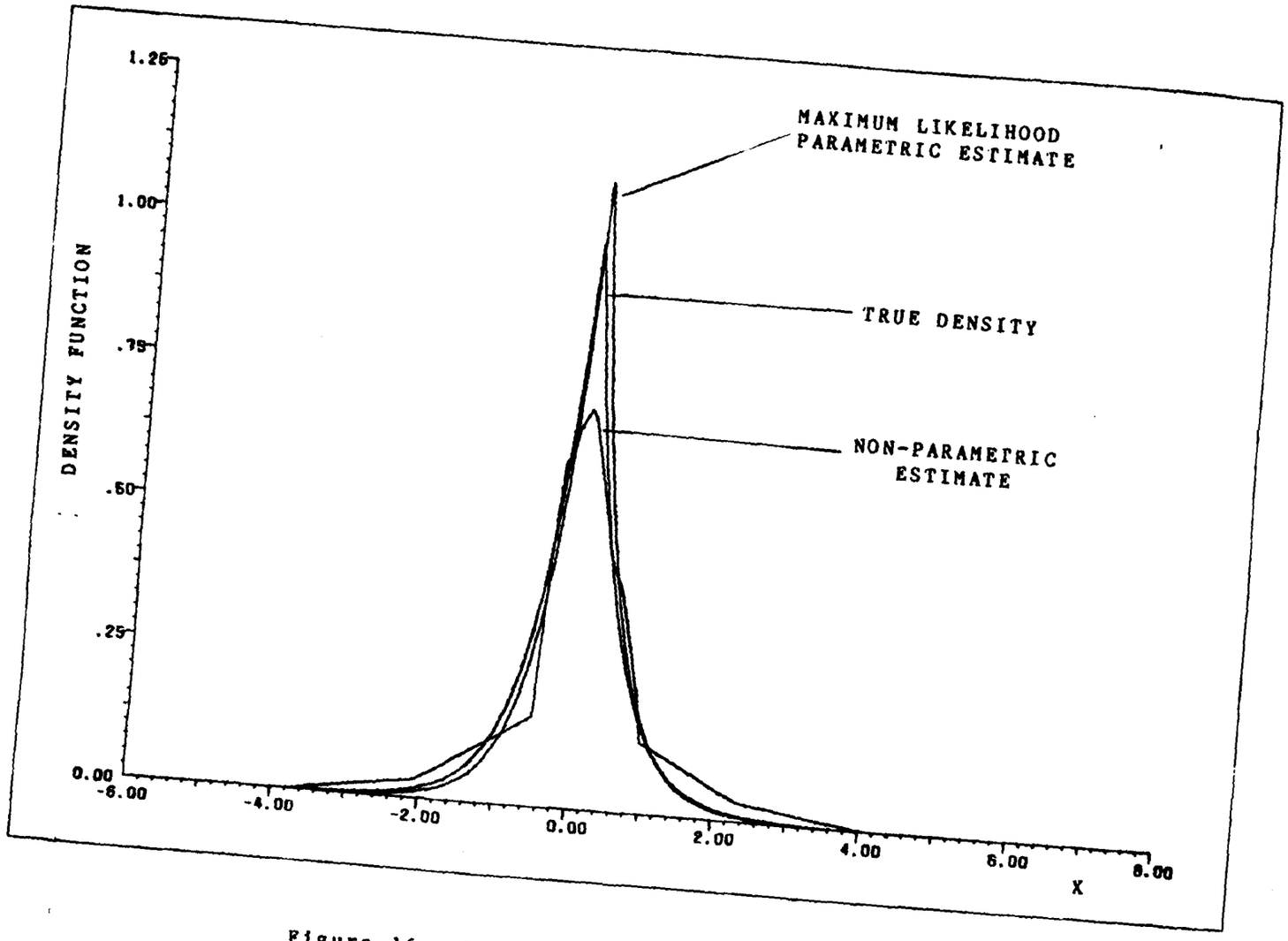


Figure 16 - Laplace Density Estimate (n=10)

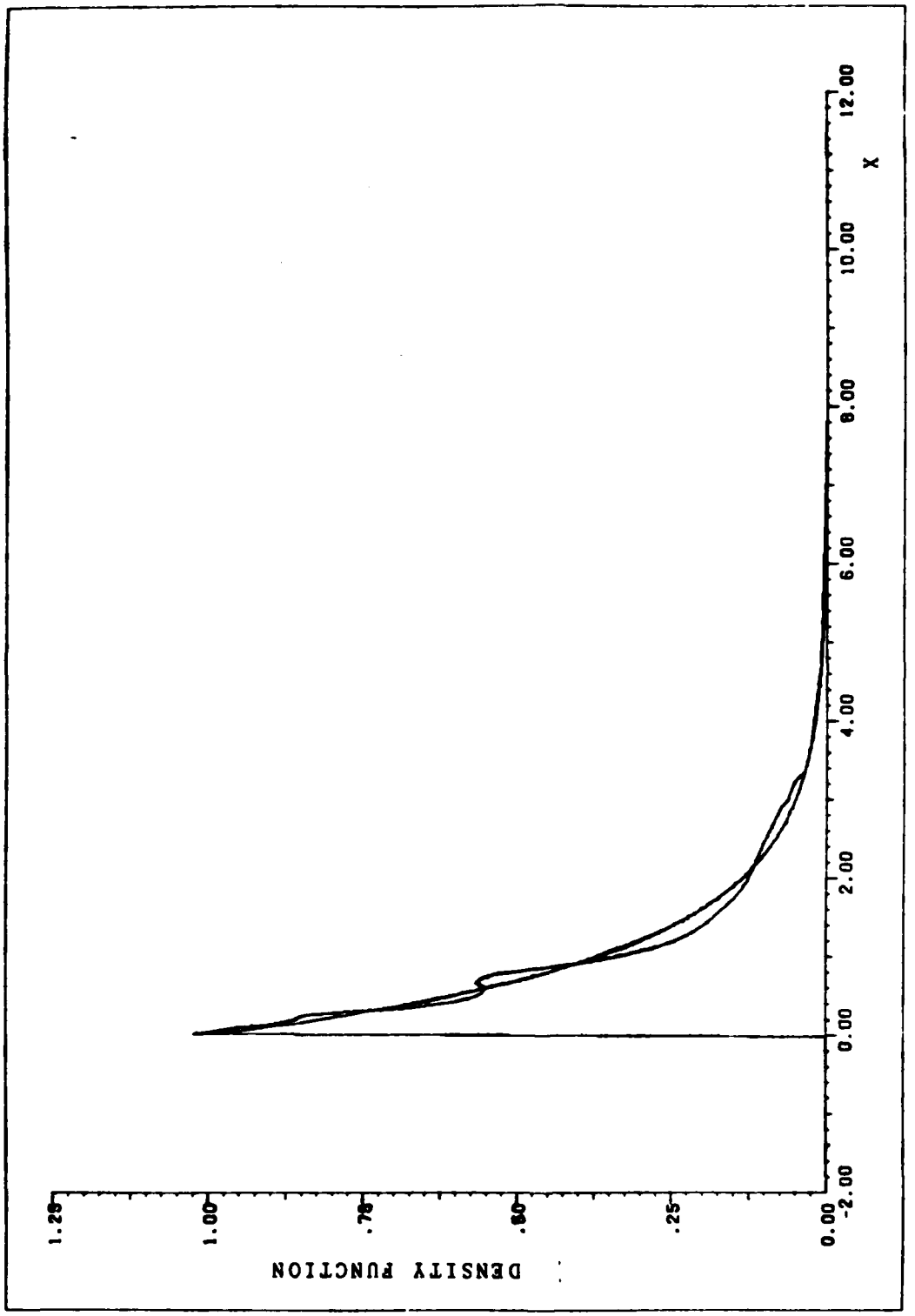


Figure 17 - True and Estimated Exponential Densities (n=100)

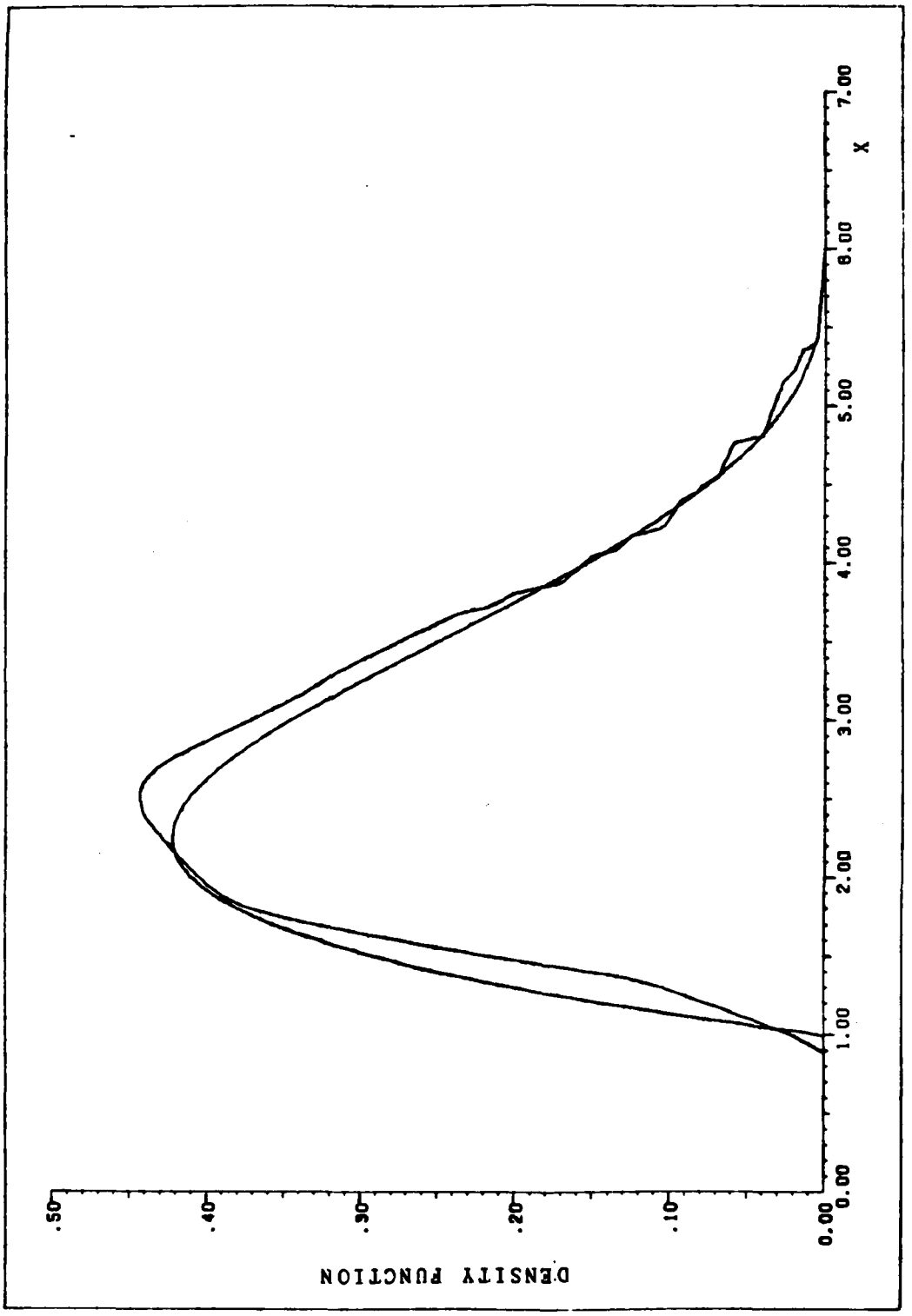


Figure 18 - True and Estimated Beta(2,4) Densities (n=100)

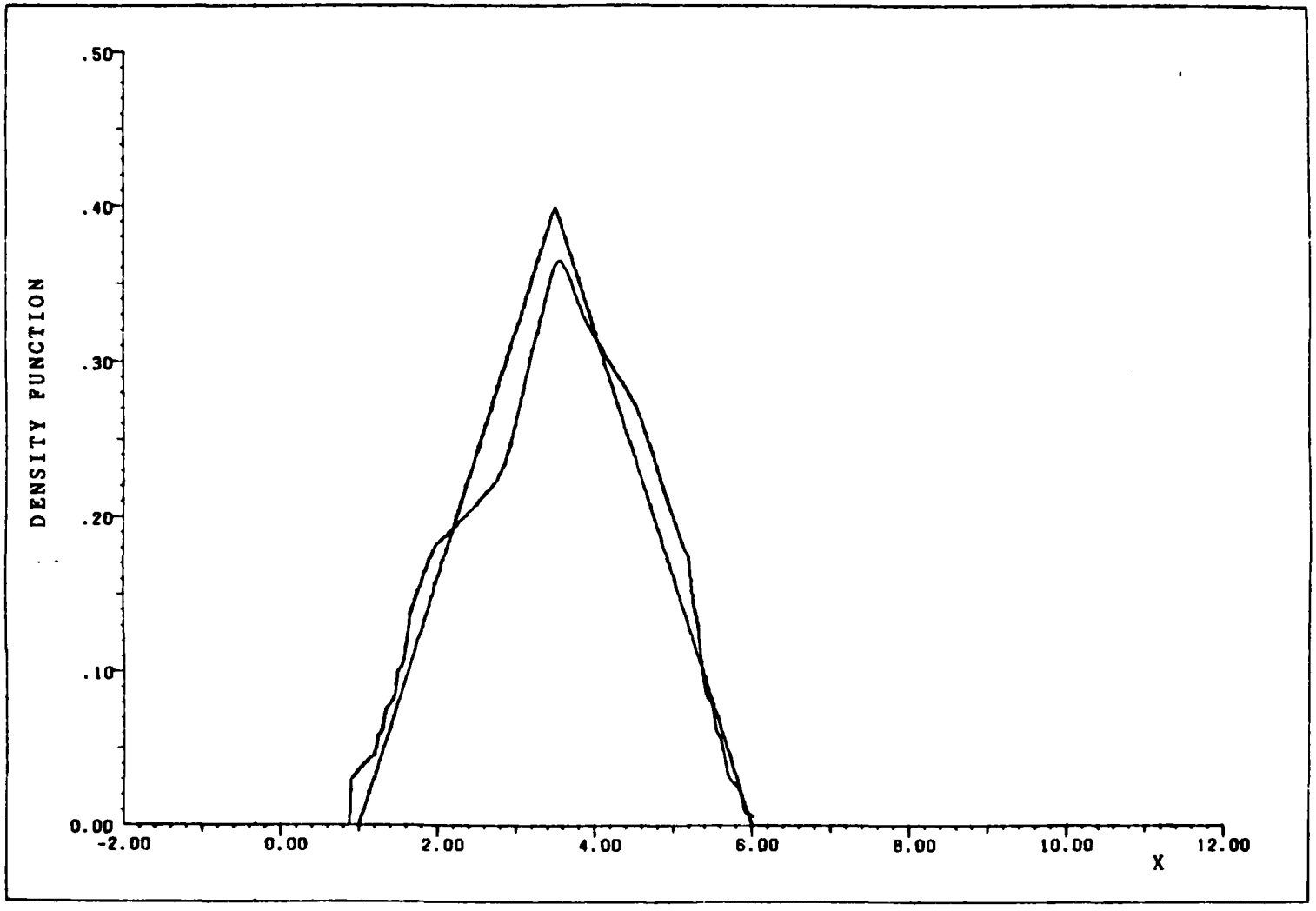


Figure 19 - True and Estimated Triangular Densities (n=100)

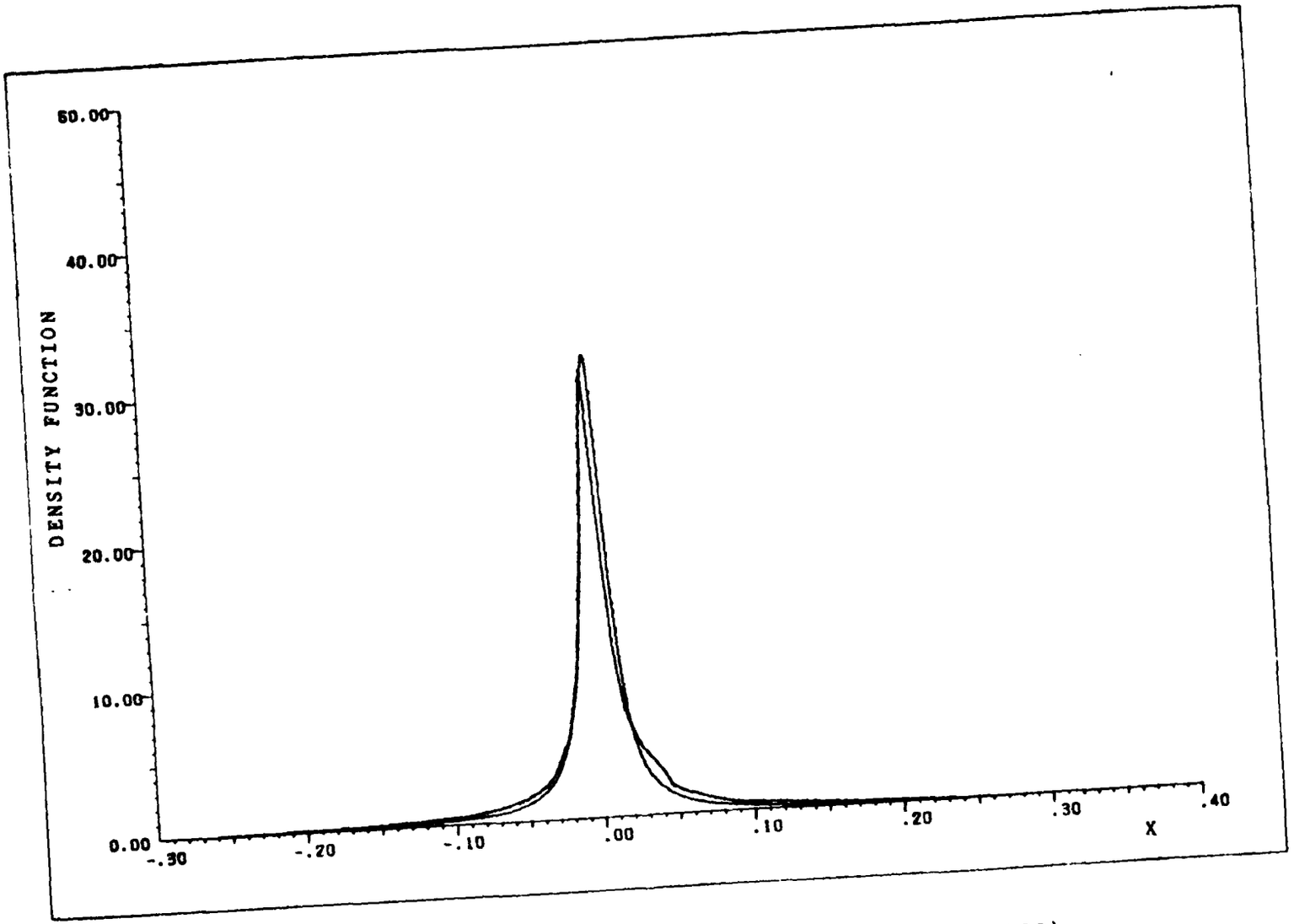


Figure 20 - True and Estimated Cauchy Densities (n=100)

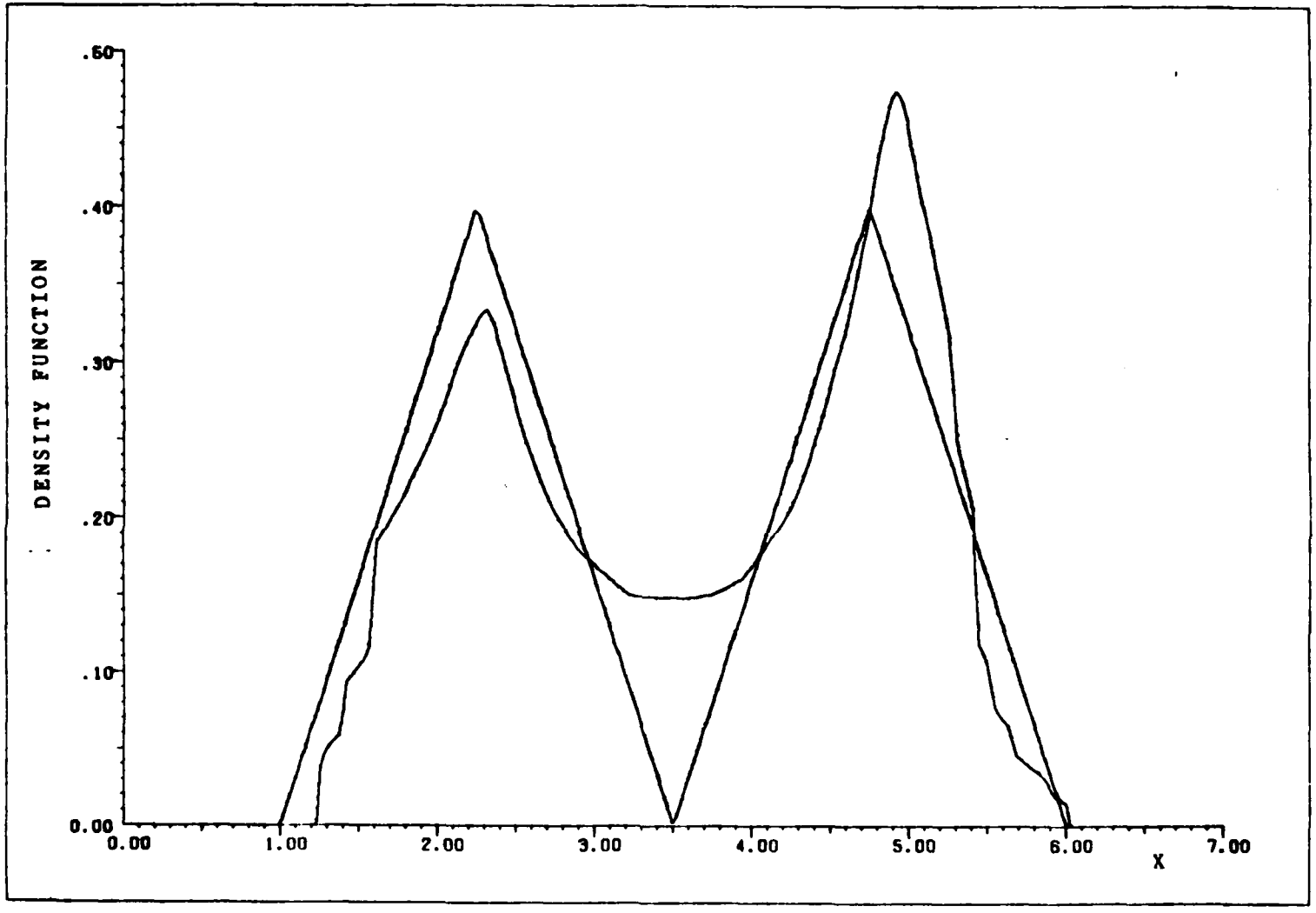


Figure 21 - True and Estimated Double Triangular Densities (n=100)

function is possible if one does not worry about the density function. During the development of this estimator, it appeared that fewer smoothing operations would improve the quality of the distribution function estimate slightly.

Figures 22 to 24 show the actual, estimated and empirical distribution functions corresponding to the estimates shown in Figures 14, 15 and 16 respectively. The improvement over the empirical distribution function is clearly evident in these plots. Also notice the smoothness of the estimated distribution function.

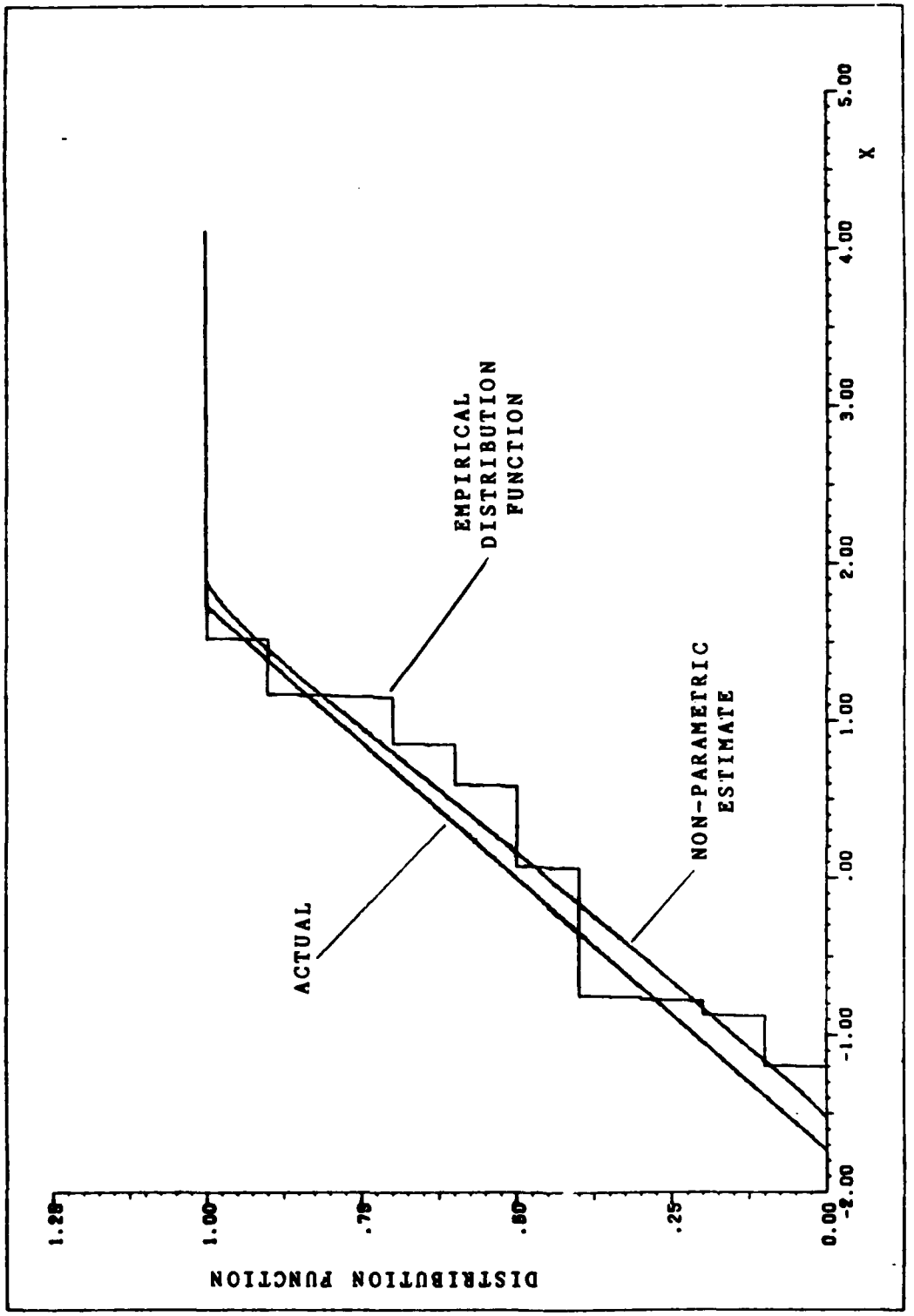


Figure 22 - Distribution Function Estimates for Uniform (n=10)

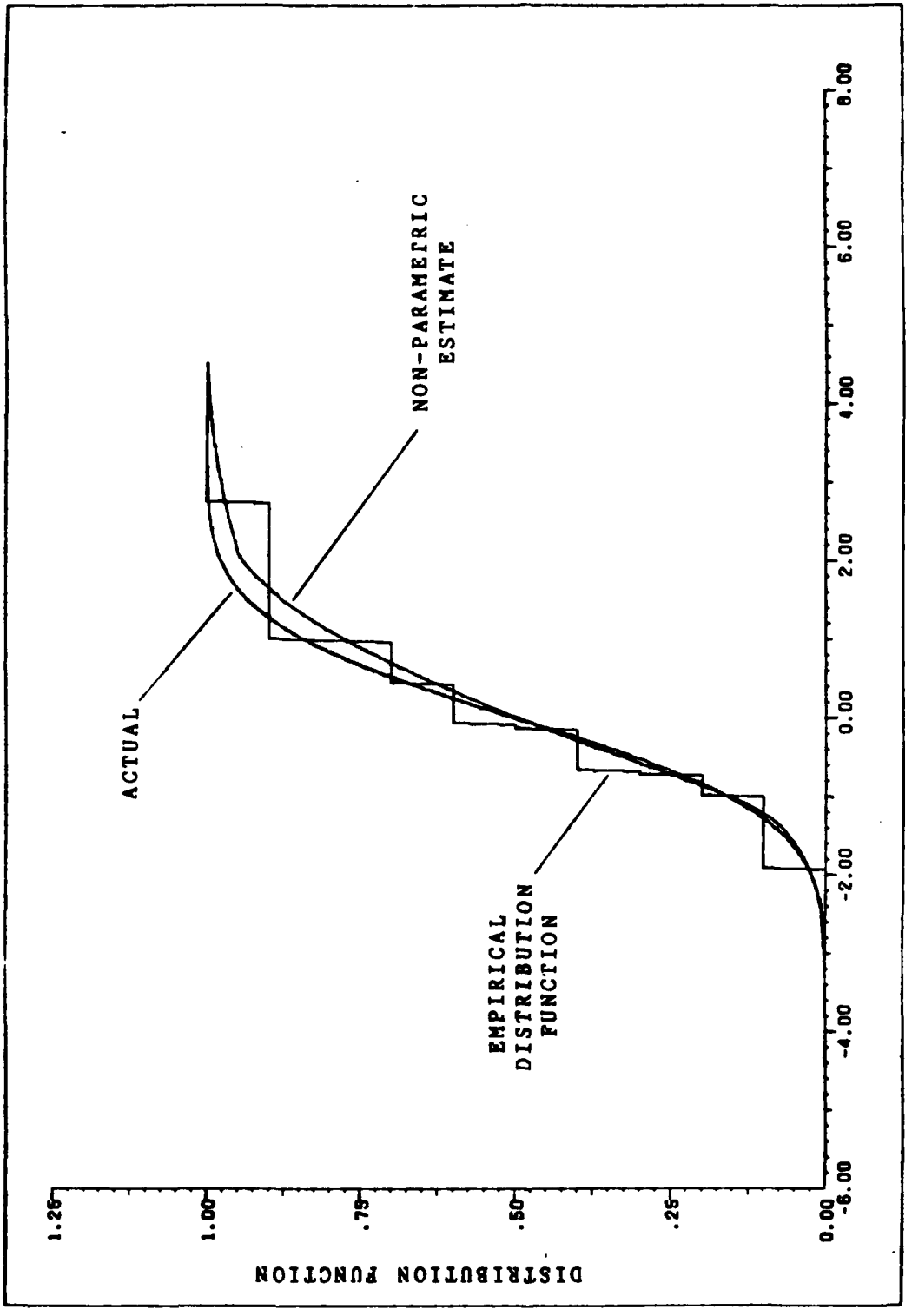


Figure 23 - Distribution Function Estimates for Normal (n=10)

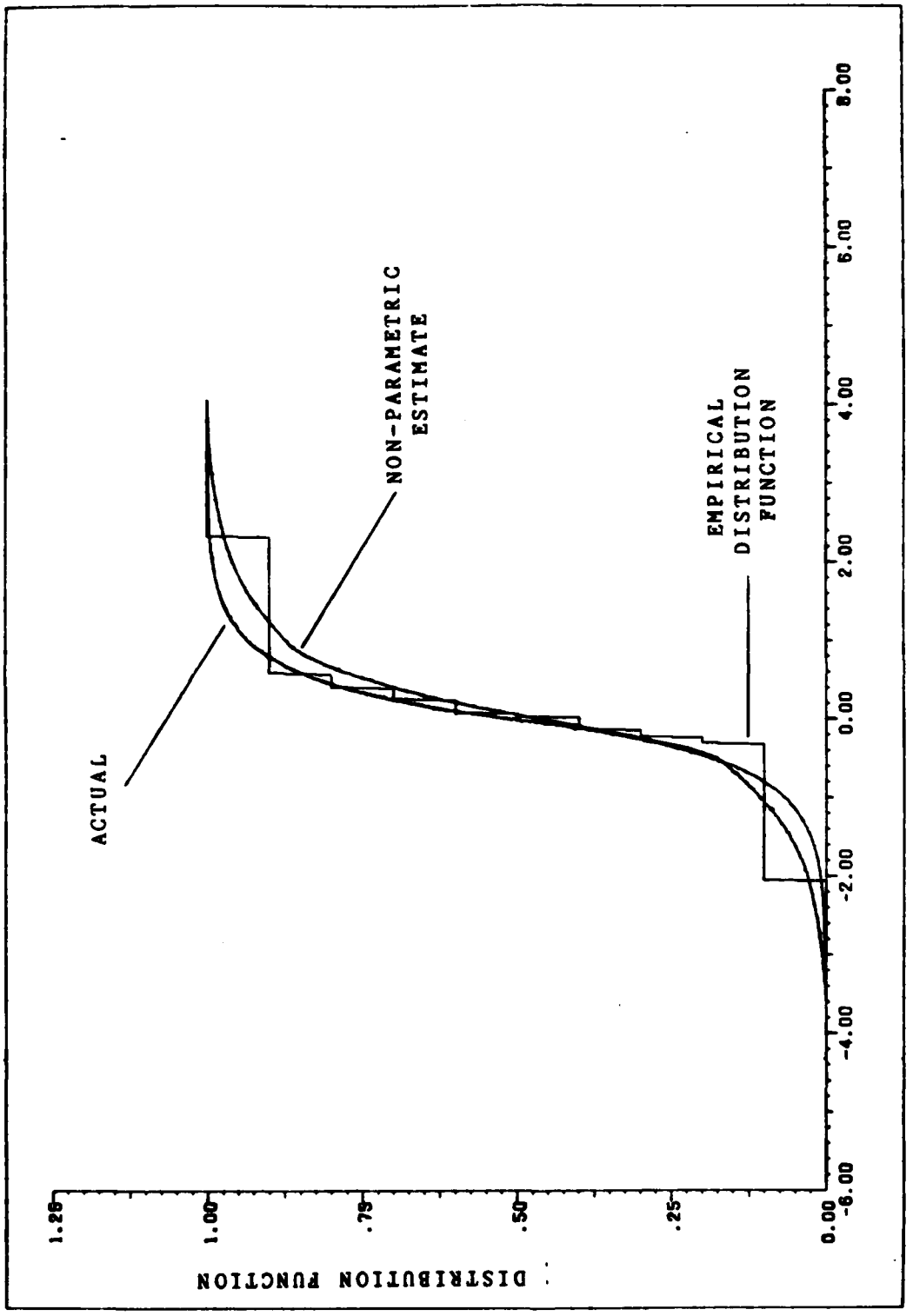


Figure 24 - Distribution Function Estimates for Laplace (n=10)

IV. Applications.

This section deals with ways of applying the new estimator derived in the previous chapter. The first application, distance estimation, is included here because one frequently wishes a simpler formulation for the density function in order to perform some other application. The small sample analysis section demonstrates the use of the new distribution function to perform a task which is poorly accomplished by the more traditional distribution function. The two sample test uses the estimated density directly to improve the power of an existing test. Many more applications are possible, and a few of these are discussed in the section on other applications.

IV.1 Distance Estimation

One of the problems with non-parametric estimators is the difficulty in using the resultant estimate. This is due to the form of the estimator and the number of "parameters" required to specify which of the members of that form is the estimate. For instance, in order to completely specify the proposed density estimate at least 2n+4 "parameters" and possibly as many as 3n (depending on subsample endpoints) are required. It would be useful to reduce the number of parameters by converting to another functional form. A logical approach to this is to try to

convert the non-parametric estimate to one member of a family of parametric distributions. This has been accomplished through minimum distance estimation.

In general, the minimum distance estimate is a set of parameter values minimizing some distance function defined in terms of the hypothesized probability structure and the sample generated estimate. Originally developed by Wolfowitz (234), minimum distance estimation attempts to match the entire probability structure (rather than the moment structure) of the data with a specified model. Minimum distance estimators have been proposed by Beran (12), Parr (135), Daniels (35) and others (See 136) and have been shown to possess some significant robust characteristics. However, most of the Monte Carlo investigations of these estimators have been confined to distributions close to the normal.

A series of logical candidates for the distance estimation task was considered. These include:

1) General Exponential Power Distribution. This distribution models symmetric data which can be extremely leptokurtic or platykurtic. Extremes of the distribution occur at the uniform model, where p = infinity, and at the double exponential, where p = 1. Moderate tail lengths are a function of the shape parameter, p. For p = 2, the distribution becomes Normal. The probability density function is:

$$f(x;p,\mu,\sigma) = [pg(p)/2\Gamma(1/p)\sigma]\exp[-|g(p)(x-\mu)/\sigma|^p]$$

$$g(p) = [\Gamma(3/p)/\Gamma(1/p)]^{.5}$$

$$-\infty < x,\mu < \infty \qquad 0 < \sigma < \infty \qquad 1 \leq p \leq \infty$$

2) Generalized Beta Distribution. This distribution models a wide variety of shapes over a bounded interval, [a,b]. Symmetric, asymmetric, and U-shaped distributions are possible. The uniform model is a special case. The wide variety of shapes and finite support (which always occurs for "real life" data) make this family especially attractive. The fact that common densities such as the Normal, Exponential, etc. are not included in the family is a detriment. The probability density function is:

$$f(x;p,q,a,b) = (x-a)^{p-1}(b-x)^{q-1}/[B(p,q)(b-a)^{p+q-1}]$$

$$B(p,q) = \Gamma(p)\Gamma(q)/\Gamma(p+q)$$

$$a \leq x \leq b \qquad 0 < p < \infty \qquad 0 < q < \infty$$

3) Generalized Gamma Distribution. This distribution models asymmetric data on the half open interval [c,). An important distribution in life testing, the generalized Gamma includes the negative exponential as a special case. The density function is:

$$f(x;a,b,c,p) = p(x-c)^{bp-1}\exp\{-[(x-c)/a]^p\}/a^{bp}\Gamma(b)$$

$$0 \leq c \leq x < \infty \qquad 0 < a,b,p < \infty$$

4) Generalized t Distribution. This distribution models symmetric data with moderate to extremely leptokurtic distributions. As the degrees of freedom parameter, n, approaches infinity, the distribution approaches the normal. For n=1, the distribution reduces to the Cauchy. The general density function is:

$$f(x;\mu,\sigma,n) = [(n+1)/2][1+(x-\mu)^2/\sigma^2 n]^{-(n+1)/2}/\Gamma(n/2)\sigma(\pi n)^{.5}$$

$$-\infty < x,\mu < \qquad 0 < \sigma < \infty \qquad 1 \leq n < \infty$$

5) R-S Distribution. This distribution was originally developed by Ramberg and Schmeister (155) to generate random variates. It is a generalization of Tukey's lambda function (205) and can be used to model a wide variety of data shapes. The probability density function is given in terms of the percentile function, R(p).

$$f(x;p,a,b,c,d) = f(R(p)) = (cp^{c-1}+d(1-p)^{d-1})/b$$

$$R(p) = a+(p^c-(1-p)^d)/b$$

$$-\infty < a \leq x < \infty \qquad -\infty < b,c,d < \infty \qquad 0 \leq p \leq 1$$

6) Generalized Life Model. Developed by Moore and Bilikam, this model includes as special cases the Weibull and the Raleigh distributions. The probability function is given by:

$$f(x;a,b,g(x)) = bg'(x)(g(x))^{b-1}\exp[-(g(x))^b/a]/a$$

$$g(x) \in R \qquad \lim_{x\to 0^+} g(x) = 0 \qquad \lim_{x\to\infty} g(x) = \infty$$

$$g(x) \text{ strictly increasing} \qquad 0 < x,a,b < \infty$$

The Generalized Beta Distribution was chosen as the family to consider for parameterizing the estimate. The distance measures considered were those discussed in Chapter III, and approximate MISE was chosen as an acceptable measure based upon tests of ten samples and the variability of the measure. Both the estimated probability density function and estimated cumulative distribution function were evaluated in the minimization scheme, but the probability density function based method was eliminated due to convergence problems in the optimization procedure and the frequency of local minima.

The distance measure was minimized using the routine ZXMIN on the International Mathematical and Statistical Libraries. For each sample three starting points were used which were chosen as the minimum distances out of a grid of eighty-one points. The Beta parameters were bracketed by choosing the largest spike in the non-parametric estimate of the density and adjusting the sizes of p_{max} and q_{max} by the ratio of the number of points to the left of the average of $x_{(1)}$ and $x_{(n)}$ to the number of points to the right of the average.

$$R = \frac{\text{Points to Left of Average}}{\text{Points to Right of Average}}$$

$$B = \max \frac{j/(n+1)}{x_{(i)} - x_{(i-j)}}$$

$$i = 1,2,\ldots,n \qquad i-j = 1,2,\ldots,n$$

Then

$$p_{max} = \max(3, B/R)$$

$$q_{max} = \max(3, BR)$$

The points for evaluating the distance function were then chosen as:

$$p_i = ip_{max}/10 \qquad i = 1,2,\ldots,9$$

$$q_i = iq_{max}/10 \qquad i = 1,2,\ldots,9$$

The distance measures were calculated at each of the eighty-one starting points and the best three points were chosen for starting a modified gradient search. If two of the three starting points did not converge to the same Beta parameters, additional starting points were chosen. If the parameters converged to some point outside p_{max} or q_{max} then additional points were checked to investigate the possibility of a local minimum.

Table 6 - Average square error for basic and parameterized estimates, n=100

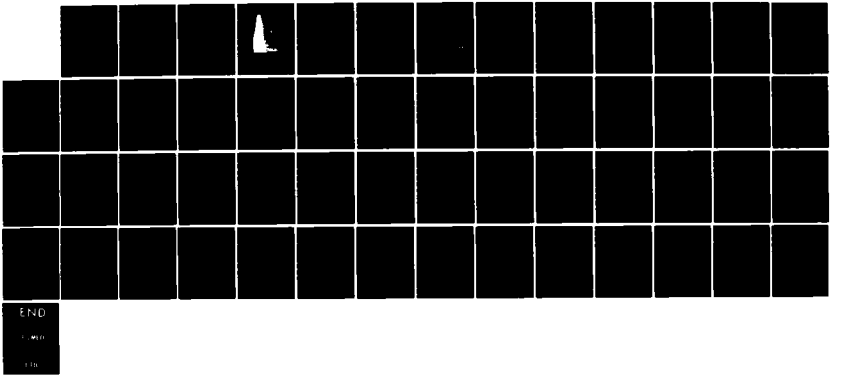
Distribution	Average Square Error			
	pdf		CDF	
	Basic	Beta fit	Basic	Beta fit
Cauchy	.0143	.1135	.00270	.00301
Laplace	.0059	.0355	.00151	.00906
Normal	.0012	.0003	.00054	.00022
Uniform	.0048	.0041	.00104	.00042
Beta(.6,.8)	.0061	.0020	.00181	.00127
Beta(2,3)	.0010	.0006	.00054	.00038

Results of the distance estimation using a beta distribution are shown in Table 6. For beta family and distributions with lighter tails than the normal the method resulted in acceptable estimates. For distributions with heavy tails, double exponential, Cauchy, etc., the method results in a poor fit. This is reasonable since the beta distribution has lighter tails than the normal even as the support becomes very large.

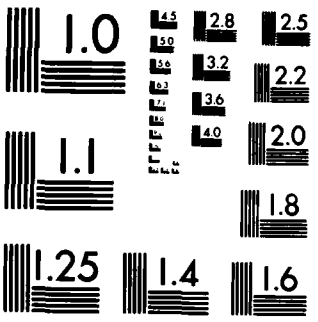
The distance estimation results are based upon medians of twenty-five minimum distance fits to each of the distributions. The small number of runs was a result of difficulties in getting the minimization procedure to converge and the large number of local minima, particularly with the more leptokurtic distributions.

AD-A151 853 A NON-PARAMETRIC PROBABILITY DENSITY ESTIMATOR AND SOME APPLICATIONS(U) AIR FORCE INST OF TECH WRIGHT-PATTERSON AFB OH SCHOOL OF ENGINEERING R P FUCHS MAY 84 2/2

UNCLASSIFIED AFIT/DS/ENC/84-1 F/G 12/1 NL



END



IV.2 Application to Small Sample Analysis

There are many instances where the cost of obtaining a single sample point of a random variable is extremely high. This is particularly true in instances of testing or analyzing complex physical systems, for example aircraft stress analysis. In these cases, one is sometimes asked to make an estimate based upon a mere handful of data points. The estimator we have developed can be used as a variance reduction technique in cases such as these.

As a theoretically interesting example, consider the determination of π by a Monte Carlo technique. If we consider a circle inscribed within a unit square as in Figure 25 and generate uniformly distributed random variable pairs (x,y) on [0,1] x [0,1], then the

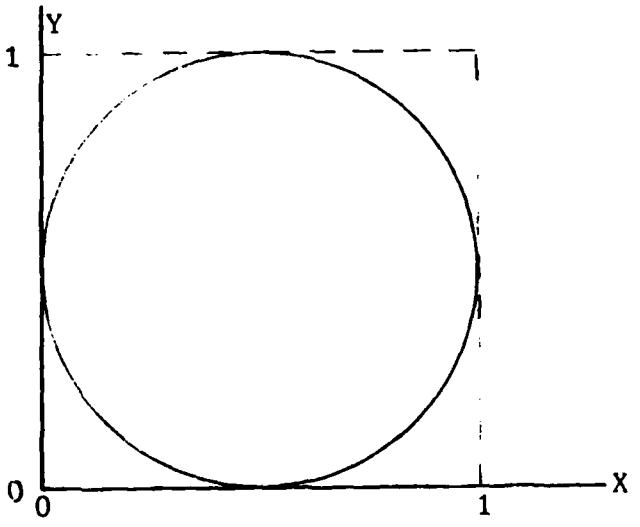


Figure 25 - Geometry for Calculation of π

probability of any (x,y) pair lying within the circle is equal to the ratio of the circle's area to the square's area or $\pi/4$. If we wish to calculate an estimate of π we may do so by taking the ratio of hits, (x,y) pairs within the circle, to total pairs generated and multiply by 4.

$$= \frac{\text{Pairs in Circle}}{\text{Total Pairs}} \times 4$$

Another way of looking at this problem is to define a new random variable

$$Z = X^2 + Y^2$$

then:

$$\hat{\pi} = 4\hat{P}(z \leq 1) = 4\hat{F}_z(1)$$

If we calculate the distribution of z we may estimate π in this manner.

Figure 26 shows the results of estimating π by the Monte Carlo method and by the distribution function method. The curves shown for the distribution function method are based upon Monte Carlo analysis with one thousand repetitions at each sample size (2,4,5,10,20, 40,100). The curves for the Monte Carlo method are determined using binomial probabilities. In addition, the percentage of times the distribution function method beat the Monte Carlo method was calculated. Table 7 presents these results.

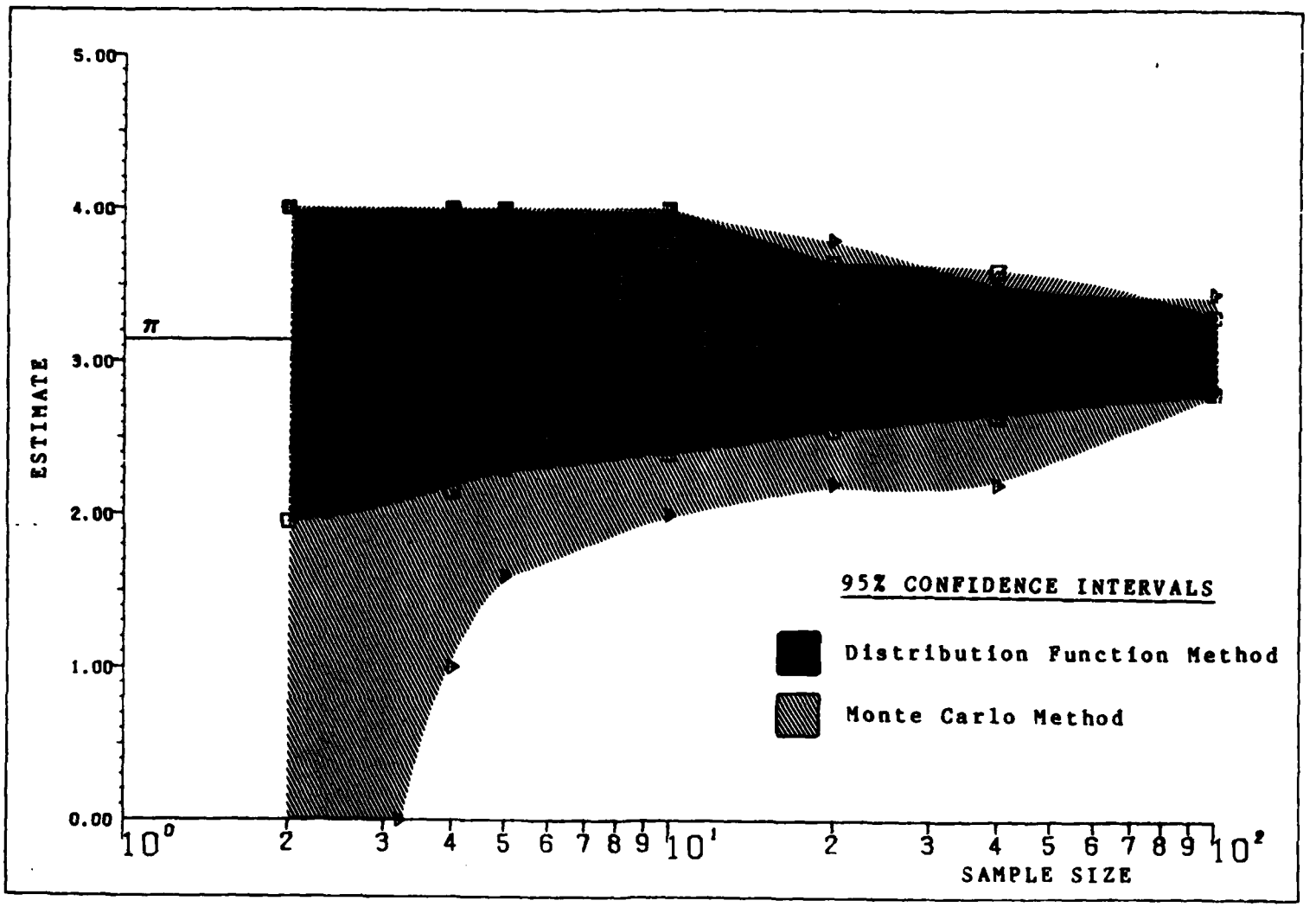


Figure 26 - Confidence Intervals for Small Sample Technique

This analysis would tend to indicate that for this type of problem one would wish to use the distribution function method for sample sizes less than about fifty rather than inferring information directly from the sample itself. This is certainly true when sample sizes get down to about five. Note that if a larger number of samples is available the Monte Carlo method is preferable when a single point in the distribution is sought.

Table 7 - Distribution Function Method Compared to Monte Carlo Method.

Sample Size	Percent Equal to or Better than Monte Carlo
2	97
4	82
5	72
10	58
20	54
40	54
100	41

IV.3 Percentage Point Estimation

A natural extension of the case presented in the last section is the estimation of a percentage point of a distribution. Figures 27 and 28 shows the 95% confidence intervals for the error in estimating percentage points of

the uniform, normal, and double exponential distributions.

The errors shown are actual squared error calculated from 100 random samples of sizes ten and 100 from each of the distributions. The horizontal axis data is first scaled to make $F^{-1}(.99) - F^{-1}(.01) = 1$ for each distribution. Then the plotted values represent:

$$[F^{-1}(i/100)-\hat{F}^{-1}(i/100)]^2 \qquad i=1,2,\ldots,99$$

Ninety-five percent of the errors are less than or equal to the curves shown. In all cases, the median error was approximately one order of magnitude smaller than the 95% confidence bound.

Although the errors appear to decrease as one estimates a point farther out in the tail of a distribution, this is not entirely accurate. For distributions with infinite support the errors increase again as a smaller percentage point is estimated. This is due to the inherent finite support of the estimator. A reasonable guideline would be that the sample size should be, as a minimum, approximately equal to the inverse of the desired percentage point. That is, if we wish to estimate $x_{.001}$ we should start with a sample size of about 1000 or greater points. The increase of accuracy near the extreme percentage points apparent in Figures 27 and 28 is due primarily to the accuracy of the endpoint estimate.

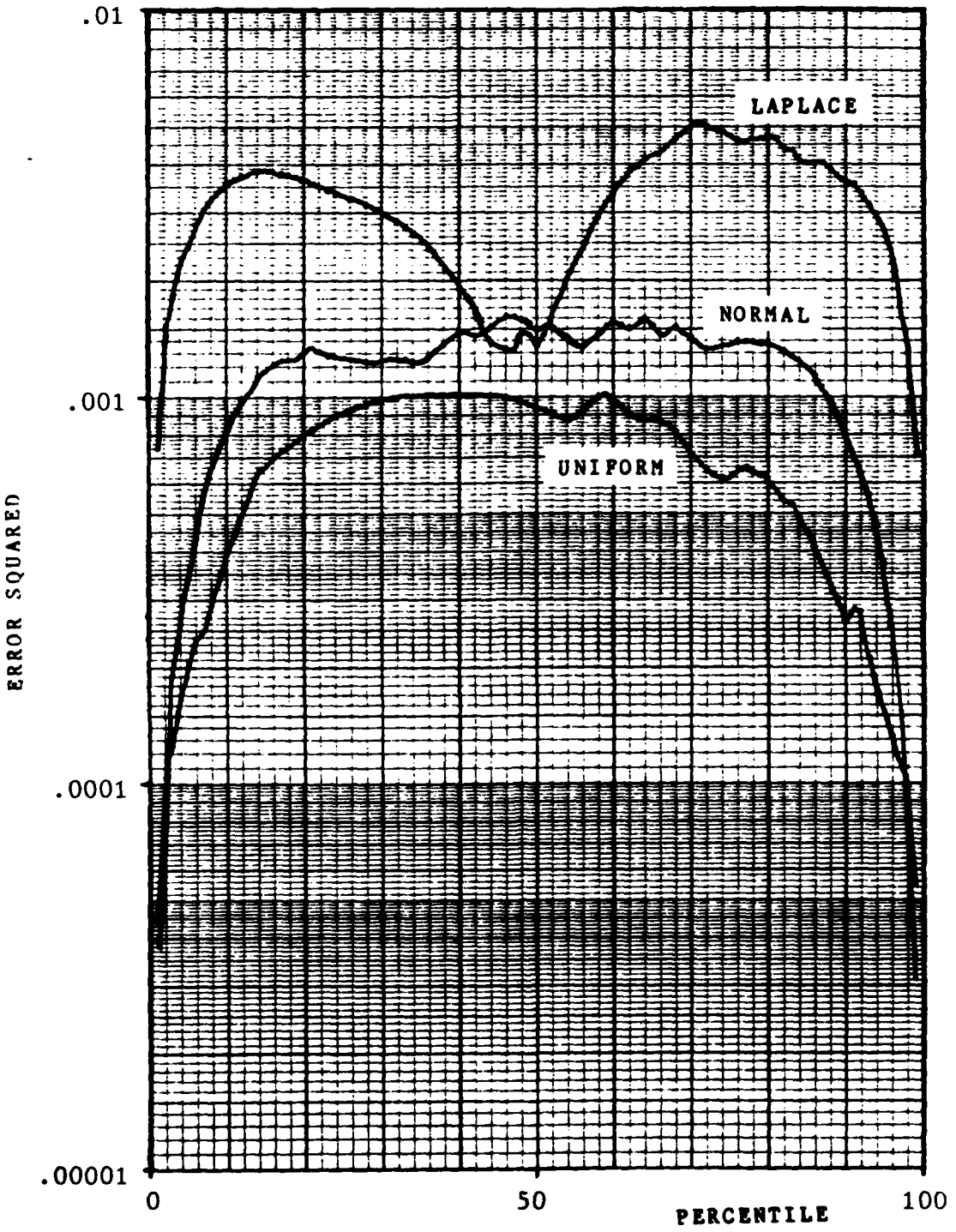


Figure 27 - 95% Upper Confidence Bound on Errors at Various Percentage Points (n=100)

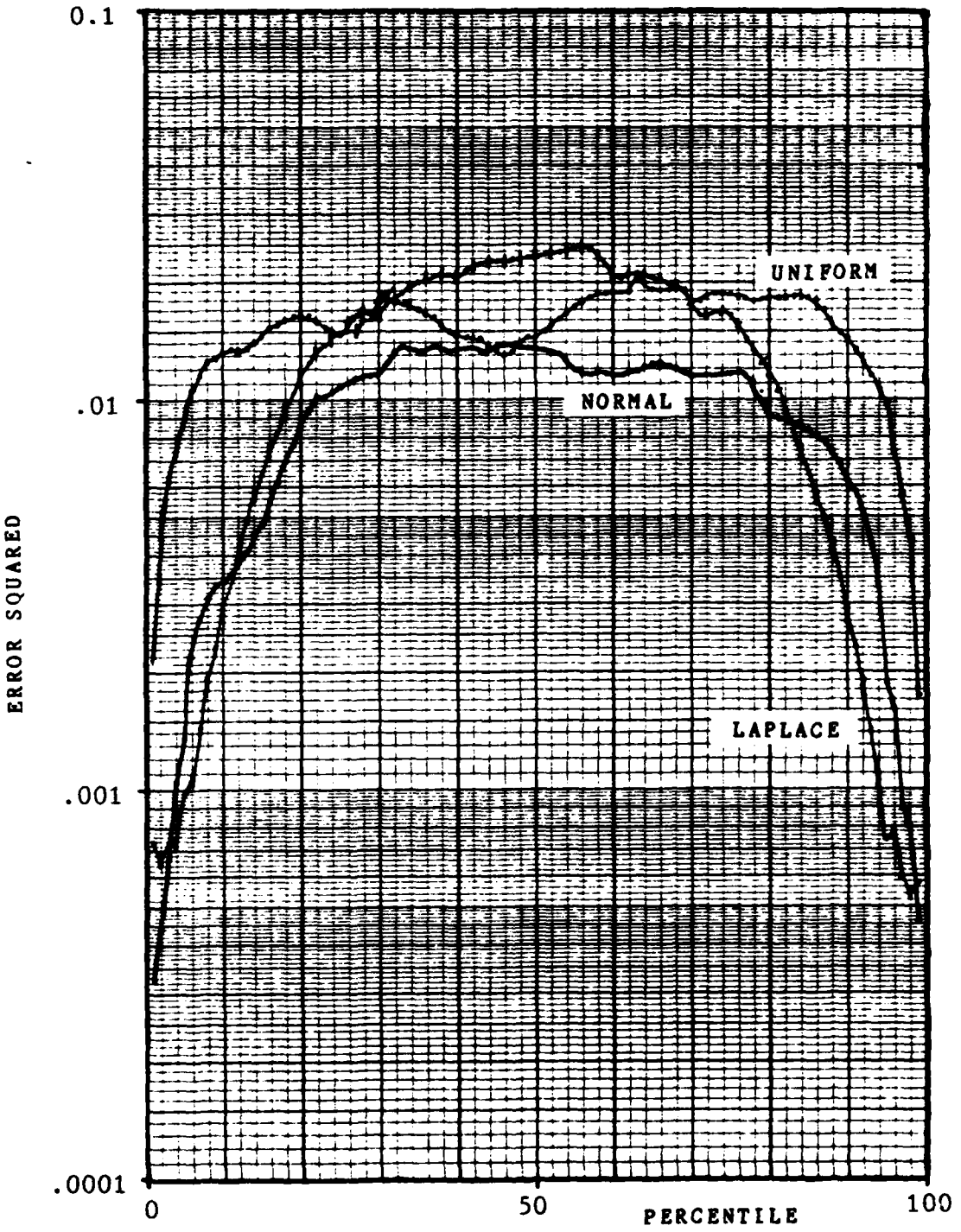


Figure 28 - 95% Upper Confidence Bound on Errors at Various Percentage Points (n=10)

IV.4 Two Sample Test

The classical two-sample tests take two random samples from independent distributions and test for some similarity (equivalence, same location, same scale) in the underlying distributions based on the sample. The usual approach is to combine the two samples into one and use as a test statistic some function of the ordering of the combined sample. Some popular tests of this genre are the Wald-Wolfowitz test (218), the Smirnov test (190), the Cramer-von Mises test (56), the median test (56), and the Mann-Whitney test (117). Many other two-sample tests exist, but none use the samples independently to evaluate a probability density function and then use the resultant density estimates in testing the null hypothesis. There are two good reasons for this. The first is that theoretical distributions of the test statistic are virtually impossible to derive due to the complexity. The second is that the noisy properties of a density function estimate can tend to obscure the true underlying density unless the estimator is very good indeed. The first problem can be somewhat overcome by using Monte Carlo methods to develop the critical values of the test statistic. The second has not, to this author's knowledge, been previously successfully attempted.

The ability to generate a density function which accurately represents the sample's underlying density

suggests that a test of the hypothesis that two samples come from the same distribution might be more powerful if performed upon the densities rather than on distribution functions. Power studies on two-sample tests are not generally available in the literature, mostly due to the difficulty in picking one of the infinite alternative distributions to test against. The two-sample Smirnov Test is generally felt to be as powerful as any (56) of the popular two-sample tests and will be used for comparison purposes in the development of a new two-sample test.

The uniform, normal, beta (1,3), beta (.5,.5), and Laplace distributions, all with zero mean and unit variance, were used as representative distributions in the development of the critical values. One hundred estimates were made from samples of sizes ten and one hundred from each distribution. The test statistic used was the integrated absolute error between all possible combinations of two estimates. This resulted in 4950 samples of the test statistic for each distribution, a total of 24750 points.

The null hypothesis to be tested was:

$$H_0 : \quad F_0 = F_1 \quad \text{or} \quad f_0 = f_1$$

against the two-sided alternative hypothesis:

$$H_a : \quad F_0 \neq F_1 \quad \text{or} \quad f_0 \neq f_1$$

Tests using both the cumulative distribution function and the probability density function were developed and compared to the Smirnov test applied to the same samples.

The critical values of the test statistic are given in Table 8. The cumulative distribution function test

Table 8 - Critical Values of the Two-Sample Test Statistic

Significance Level	CDF Test		pdf Test	
	n = 10	n = 100	n = 10	n = 100
.001	10.258	.07861	1.3814	.01130
.005	10.185	.07707	1.1618	.01058
.01	10.043	.07601	1.1294	.00988
.05	4.1517	.03263	.6989	.00711
.10	2.6214	.02067	.4915	.00484
.15	2.1287	.01339	.3467	.00436
.20	1.7304	.01108	.2975	.00400
.25	1.4039	.00953	.2491	.00375
.30	1.0826	.00792	.2041	.00352
.35	.8904	.00654	.1179	.00336
.40	.7795	.00536	.1609	.00319
.45	.6610	.00467	.1454	.00305
.50	.5487	.00410	.1291	.00292

statistics are calculated using the same results as those used in the pdf critical value computations. They are given primarily to allow a direct power comparison with the Smirnov test for this particular set of samples.

The critical values in Table 8 were used in a power study with all possible combinations of the chosen samples. This resulted in 10000 samples from each pair of dissimilar distributions. The power of the test was calculated by taking the proportion of samples correctly rejected as failing to meet the hypothesis. Tables 9, 10, 11, 12, 13, and 14 show a comparison of the powers of the three tests.

The power studies indicate that:

1) The test based on the new probability density function is consistently more powerful than the cumulative distribution function based test developed here.

2) The Smirnov test is slightly more powerful than the probability density function test for uniform alternatives.

3) The test based on the probability density function is clearly superior for differentiating a normal from a double exponential sample.

Figures 29(a) and (b) show the distribution and density functions of the distributions used in the power studies for this test. One could argue that the normal and double exponential are more clearly distinquishable in

α - Level	Laplace vs.					
	Normal			Uniform		
	pdf - Test	Smirnov	CDF - Test	pdf - Test	Smirnov	CDF - Test
.001	.0002	0	0	.0011	0	0
.005	.0012	0	0	.0060	0	.0002
.01	.0015	0	0	.0077	.0075	.0010
.05	.0390	.0003	.0167	.0602	.0812	.0636
.10	.0765	.0131	.0504	.1188	.2695	.1409
.15	.1080	.0131	.0715	.2331	.2695	.2121
.20	.1323	.0590	.0915	.2933	.5697	.3237
.25	.1824	.0590	.1420	.3814	.5697	.3237
.30	.2357	.0590	.2145	.4984	.5697	.4397
.35	.2886	.0590	.2530	.5808	.5697	.5758
.40	.3369	.0590	.2816	.6430	.5697	.5758
.45	.3818	.1703	.3372	.6988	.8218	.6425
.50	.4349	.1703	.3881	.7631	.8218	.7018

Table 9 - Power Comparisons for Two-Sample Tests (n=10)

α - Level	Normal vs.					
	Uniform			Laplace		
	pdf - Test	Smirnov	CDF - Test	pdf - Test	Smirnov	CDF - Test
.001	0	.0139	0	.0002	0	0
.005	.0003	.0139	0	.0012	0	0
.01	.0005	.0315	0	.0015	0	0
.05	.0244	.1051	.0041	.0390	.0003	.0167
.10	.0584	.2665	.0435	.0765	.0131	.0504
.15	.1395	.2665	.0727	.1080	.0131	.0715
.20	.1807	.4356	.1225	.1323	.0590	.0915
.25	.2557	.4356	.1784	.1824	.0590	.1420
.30	.3530	.4356	.2603	.2357	.0590	.2145
.35	.4266	.4356	.3229	.2886	.0590	.2530
.40	.4847	.4356	.3796	.3369	.0590	.2816
.45	.5478	.7501	.4581	.3818	.1703	.3372
.50	.6162	.7051	.5512	.4349	.1703	.3881

Table 10 - Power Comparisons for Two-Sample Tests (n=10)

α - Level	Uniform vs.					
	Normal			Laplace		
	pdf - Test	Smirnov	CDF - Test	pdf - Test	Smirnov	CDF - Test
.001	0	.0139	0	.0011	0	0
.005	.0003	.0139	0	.0060	0	.0002
.01	.0005	.0315	0	.0077	.0075	.0010
.05	.0244	.1051	.0041	.0602	.0812	.0636
.10	.0584	.2665	.0435	.1188	.2695	.1409
.15	.1395	.2665	.0727	.2331	.2695	.2121
.20	.1807	.4356	.1225	.2933	.5697	.3237
.25	.2557	.4356	.1784	.3814	.5697	.3237
.30	.3530	.4356	.2603	.4984	.5697	.4397
.35	.4266	.4356	.3229	.5808	.5697	.5758
.40	.4847	.4356	.3796	.6430	.5697	.5758
.45	.5478	.7051	.4581	.6988	.8218	.6425
.50	.6162	.7051	.5512	.7631	.8218	.7018

Table 11 - Power Comparisons for Two-Sample Tests (n=10)

α - Level	Laplace vs.					
	Normal			Uniform		
	pdf - Test	Smirnov	CDF - Test	pdf - Test	Smirnov	CDF - Test
.001	.1976	0	.0006	.9992	1	.9972
.005	.2230	.0004	.0008	1	1	.9974
.01	.2507	.0014	.0008	1	1	.9978
.05	.4386	.0230	.1166	1	1	1
.10	.7345	.0584	.3135	1	1	1
.15	.8061	.0909	.5214	1	1	1
.20	.8604	.1305	.6091	1	1	1
.25	.8921	.1863	.6752	1	1	1
.30	.9180	.2543	.7521	1	1	1
.35	.9352	.2543	.8123	1	1	1
.40	.9487	.3451	.8679	1	1	1
.45	.9608	.3451	.9000	1	1	1
.50	.9691	.4511	.9248	1	1	1

Table 12 - Power Comparisons for Two-Sample Tests (n=100)

α- Level	Normal vs.					
	Laplace			Uniform		
	pdf - Test	Smirnov	CDF - Test	pdf - Test	Smirnov	CDF - Test
.001	.1976	0	.0006	.4792	1	.0322
.005	.2230	.0004	.0008	.8069	1	.0051
.01	.2507	.0014	.0008	.8556	1	.0061
.05	.4386	.0230	.1166	.9683	1	.9255
.10	.7345	.0584	.3135	.9958	1	.9941
.15	.8061	.0909	.5214	.9980	1	.9990
.20	.8604	.1305	.6091	.9990	1	.9998
.25	.8921	.1863	.6752	.9992	1	1
.30	.9180	.2543	.7521	.9993	1	1
.35	.9352	.2543	.8123	.9994	1	1
.40	.9487	.3451	.8679	.9994	1	1
.45	.9608	.3451	.9000	.9996	1	1
.50	.9691	.4511	.9248	.9998	1	1

Table 13 - Power Comparisons for Two-Sample Tests (n=100)

α- Level	Uniform vs.					
	Normal			Laplace		
	pdf - Test	Smirnov	CDF - Test	pdf - Test	Smirnov	CDF - Test
.001	.7479	1	.0032	.9992	1	.9972
.005	.8069	1	.0051	1	1	.9974
.01	.8556	1	.0061	1	1	.9978
.05	.9683	1	.9255	1	1	1
.10	.9958	1	.9941	1	1	1
.15	.9980	1	.9990	1	1	1
.20	.9990	1	.9998	1	1	1
.25	.9992	1	1	1	1	1
.30	.9993	1	1	1	1	1
.35	.9994	1	1	1	1	1
.40	.9994	1	1	1	1	1
.45	.9996	1	1	1	1	1
.50	.9998	1	1	1	1	1

Table 14 - Power Comparisons for Two-Sample Tests (n=100)

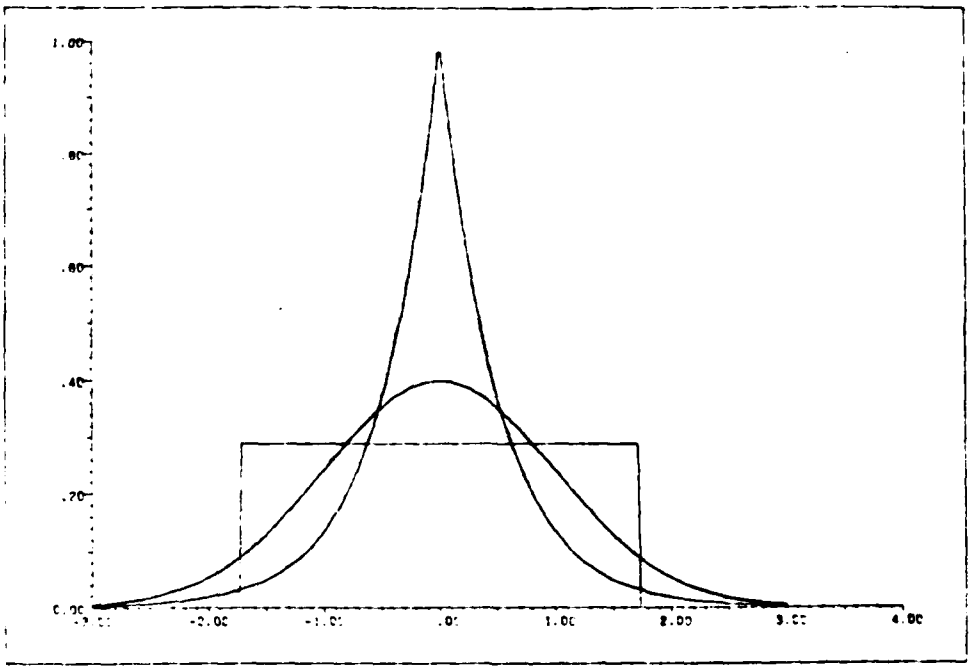


Figure 29(a) - Density Functions Used in Two Sample Tests

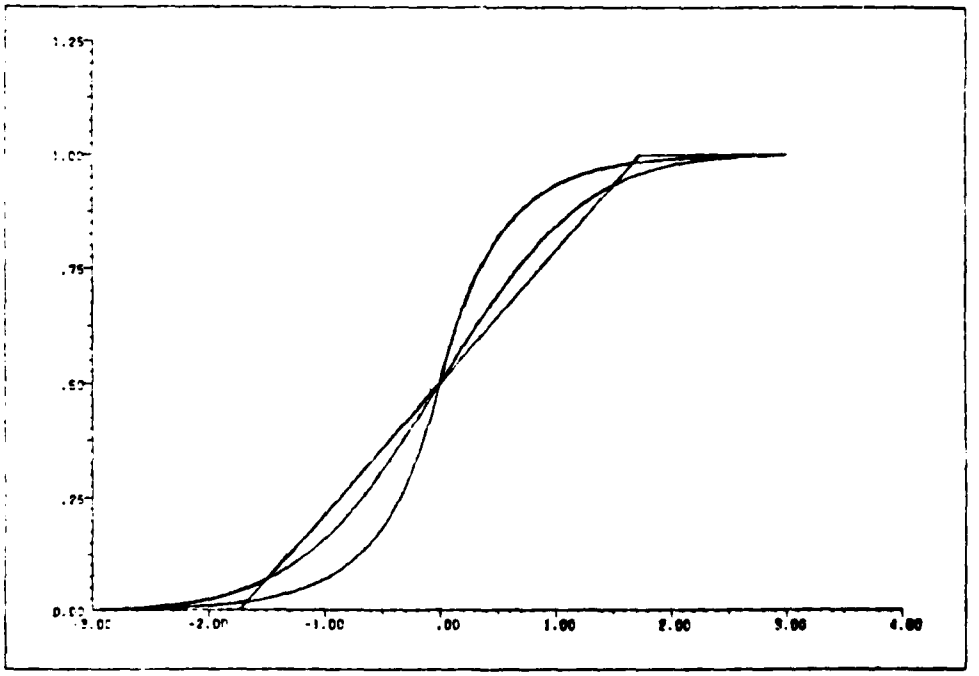


Figure 29(b) - Distribution Functions Used in Two Sample Tests

the densities. This is not as true for the uniform. This could be the reason for the improved power of the probability density function test over the Smirnov test.

The power differences might lead one to develop a two-sample test which mixes the results of both the tests discussed in this chapter to obtain an "optimal" test statistic. This "optimal" statistic would be a function of some shape statistic, such as a modified Hogg's Q or a modified kurtosis. It might also be a function of sample size, although the sample size appears to have only a small effect.

The two-sample test developed in this chapter shows clear superiority over an effective classical test under some circumstances. Additional research in this area is discussed in the Summary and Recommendations.

IV.5 Other Applications

Sample distribution functions have long been used in the areas of non-parametric goodness-of-fit tests, two sample tests, and tests for equality of some moment or other property of the distribution underlying a sample. The density function has seen little use in this area for two reasons:

1) It is difficult to calculate a estimate of the density, with the nice properties that already exist for the sample distribution function.

2) There has been no particular motivation to calculate the density.

The first reason is a valid one. In the case of the second reason there should have been some motivation, since the density estimate can provide additional information which is particularly useful in small sample situations.

In order to see the utility of the density function, and grasp the idea that further information is provided, it is instructive to consider a simple case. The Kolmogorov-Smirnov statistic (or any of several related statistics) is frequently used in various non-parametric tests. This statistic is defined as:

$$S_F = \sup_R |\hat{F}_1(x) - \hat{F}_2(x)|$$

If there were no further information to be gained from the density function then one would expect that a similar statistic defined for the density function,

$$S_f = \sup_R |\hat{f}_1(x) - \hat{f}_2(x)|$$

would be determined at the same point, x, as S_F. That is, if we consider continuous functions and define the points where the maxima occur as:

$$S_f = |\hat{f}_1(x_1) - \hat{f}_2(x_1)|$$

and

$$S_F = |\hat{F}_1(x_2) - \hat{F}_2(x_2)|$$

then $x_1 = x_2$. This is not true in general. Consider the case where $\hat{F}_1 - \hat{F}_2$ is a continuous function. Then we know that at max $|\hat{F}_1 - \hat{F}_2|$,

$$d/dx\ (\hat{F}_1 - \hat{F}_2) = \hat{f}_1 - \hat{f}_2 = 0$$

if the maximum is interior to the interval of definition. We also know that $P(\hat{F}_1 - \hat{F}_2 = 0 \quad \text{for all } x) = 0$ implies $P(S_f=0) = 0$ and $P(x_i) = 0$ for a continuous distribution and any single point (consider an endpoint), x_i, thus

$$P(x_1=x_2) = P(\hat{F}_1 = \hat{F}_2) + P(x_1=x_2=x_i) = 0$$

so the point associated with S_F is not the same as the point associated with S_f, and different information is obtained with each of the functions.

Based upon this reasoning, it seems safe to conclude that any test which has been performed with the sample distribution function can be improved by using the sample density function estimate derived here. This is not to imply that sample distribution function methods should be discarded, but rather that they should be augmented by a similar procedure using the density function estimate. The quality of such tests has already been demonstrated. Goodness-of-fit tests and multiple sample tests are logical extensions of this work.

V. Guidelines for Using the Estimator.

One always has the choice of using a parametric or non-parametric method of estimating the density. The choice is really whether or not one has enough knowledge to be able to select the distributional form a priori. The alternative is to rely upon the "gods of chance" and let the data determine the distributional form and "parameters". Using a non-parametric method is, in a sense, taking out an insurance policy. The premium will determine whether or not the policy is cost effective.

Classical estimation relies on the choice of a set of parameters from an assumed underlying probability distribution. In general, the underlying distribution is selected by extra-mathematical means and is done separately from the parameter estimation. As Fisher (52) stated in 1922, when discussing the problem of specification (as he called the selection of an underlying density function):

> "As regards problems of specification, these are entirely a matter for the practical statistician, for those cases where the qualitative nature of the hypothetical population is known do not involve any problems of this type. In other cases we may know by experience what forms are likely to be suitable, and the adequacy of our choice may be tested a posteriori. We must confine ourselves to those forms which we know how to handle, or for which any tables which may be necessary have been constructed."

Unfortunately, we are often not blessed with the insight necessary to correctly select the underlying distribution. Thus, attempts have been made to determine the form of a sample (goodness-of-fit) or to protect against incorrect assumptions by making the procedures robust (5, 8,22,23,35,53,55,66,67,72,79,81,82,83,86,109,127,135, 163,198). Traditional goodness-of-fit tests have low power for small samples. There are penalties to pay for any form of protection derived by making the estimate more robust, just as there are penalties for assuming the wrong density form and blindly applying classical procedures.

Probably the most common way to analyze data is to estimate the parameters of a distribution using an assumed probability law. Classical methods of parameter estimation abound, including the maximum likelihood method and the method of moments. Frequently, certain properties of estimators are required, such as: unbiasedness, invariance,or linearity. These further restrict the class of estimators considered and one can usually define a "best" estimator within a certain class. All of these methods assume a parametric model of the data.

An analysis of the new density estimator presented here shows one what the "premiums" are for using this estimator rather than a maximum likelihood parametric estimator. (Although the maximum likelihood estimate is not necessarily the "best", it is well known and easily

calculated for comparison purposes.) The specific estimators used were:

Uniform [a,b]:

$$\hat{a} = x_{(1)} \qquad \hat{b} = x_{(n)}$$

Normal (μ,σ^2):

$$\hat{\mu} = \bar{x} \qquad \hat{\sigma}^2 = 1/n \sum_{i=1}^{n} (x_i - \bar{x})^2$$

Double Exponential (μ,σ):

$$\hat{\mu} = \text{median}\{x_i\} \qquad \hat{\sigma}^2 = 1/n \sum_{i=1}^{n} |x_i|$$

One hundred samples of size 100 and of size ten were generated from each of the three distributions, all with zero mean and unit variance. The approximate MISE was calculated for both the cumulative distribution function and the probability density function using both estimators. The medians of the errors calculated were used to generate the tables in this section.

First the ratio of the errors:

$$R = \frac{\text{Estimator Error}}{\text{Max Likelihood Error}}$$

was calculated for each of the cases. These data are presented in Tables 15 and 16. The obvious conclusion is

Table 15 - CDF Median Error Ratios (ASE)

Sample Size	Uniform	Normal	Laplace
10	1.405	.8769	2.234
100	6.913	.8553	10.21

Table 16 - Pdf Median Error Ratios (ASE)

Sample Size	Uniform	Normal	Laplace
10	1.133	1.279	1.581
100	1.097	1.496	7.425

that for small sample sizes the penalty for using the new non-parametric estimator, even when one suspects the underlying distribution, may not be too large. It is also interesting to note that the non-parametric estimate for the normal cumulative distribution function is consistently better than the parametric maximum likelihood estimate. The non-parametric estimate resulted in lower error for 74% of the samples.

Since the entries in Tables 15 and 16 are ratios, a decision of whether or not one can accept the degradation in the estimate is dependent upon the actual error values. Tables 17 and 18 show the median error values calculated for the same samples. In general, the errors are quite

Table 17 - Median MISE for Estimates of CDF

Sample Size	Estimate	Uniform	Normal	Laplace
10	New	.00811	.00667	.00748
10	Max Lik	.00443	.00960	.00394
100	New	.00030	.00054	.00270
100	Max Lik	.00004	.00075	.00021

Table 18 - Median MISE for Estimates of Pdf

Sample Size	Estimate	Uniform	Normal	Laplace
10	New	.0135	.0101	.0261
10	Max Lik	.0103	.0072	.0146
100	New	.0011	.0011	.0066
100	Max Lik	.0008	.0008	.0008

low for many applications, thus one may be willing to pay the "premium" for using the non-parametric estimate.

The alternative is to select the underlying distribution in advance. One may potentially pay a different penalty in this case, that of selecting the wrong distribution. Tables 19, 20, 21, and 22 show the comparable errors for selecting the wrong distributions in these cases.

Table 19 - Median CDF MISE for Maximum Likelihood Estimate n = 10

		Actual Distribution		
		Uniform	Normal	Laplace
Assumed Distribution	Uniform	.00443	.02447	.03772
	Normal	.02364	.00960	.02738
	Laplace	.03311	.01999	.00394

Table 20 - Median Pdf MISE for Maximum Likelihood Estimate n = 10

		Actual Distribution		
		Uniform	Normal	Laplace
Assumed Distribution	Uniform	.0103	.1003	.1305
	Normal	.0866	.0072	.0951
	Laplace	.1121	.0888	.0146

Table 21 - Median CDF MISE for Maximum Likelihood Estimate n = 100

		Actual Distribution		
		Uniform	Normal	Laplace
Assumed Distribution	Uniform	.00004	.03333	.05805
	Normal	.03790	.00075	.00367
	Laplace	.04988	.00471	.00021

Table 22 - Median Pdf MISE for Maximum Likelihood Estimate n = 100

		Actual Distribution		
		Uniform	Normal	Laplace
Assumed Distribution	Uniform	.0008	.0588	.0900
	Normal	.0421	.0008	.0077
	Laplace	.0847	.0079	.0008

One may be tempted to say that the Q discriminant probabilities in Tables 1 and 2 coupled with the errors in Tables 19 through 22 could lead to an adaptive parametric estimator with smaller expected error than the non-parametric estimator. This is true; however, one is seldom, if ever, in the situation where a selection between only these distributions needs to be made. In the real world a selection must be made from a continuous space of distributions. The parametric estimator will pay a penalty based upon the distance between the assumed distribution and the actual distribution. The non-parametric estimator is more likely to pay a penalty related to some measure of the shape of the underlying distribution.

The question of what estimator to use ultimately depends its specific use. However, for many cases where there is uncertainty about the underlying distribution, and where the sample size is small, the risk in using the

non-parametric procedure outlined in this paper is smaller than the risk associated with using a maximum likelihood estimate.

VI. Summary and Recommendations.

A new non-parametric density estimate has been developed which has the following properties:

1) It is continuous and piecewise linear.

2) It converges to the true density function if the true density has no more than a finite number of discontinuities of a form where the value at the discontinuity can be considered the average of the limiting values on either side of the discontinuity.

3) It requires no user supplied parameters.

The estimator is shown to have significantly better error properties, for certain classes of distributions, than existing density estimators. The quality of the estimate is discussed, tabulated and graphically demonstrated. Applications, including parameterization, small sample analysis, and two sample tests are presented. These newly developed applications are shown to improve upon the generally accepted existing techniques. Guidelines for choosing a density estimation method along with a discussion of an approach to method selection are presented.

Research opportunities in the field of density estimation and its applications have been expanded by this research. In particular, the applications shown have demonstrated the utility, versatility, and strengths of

density based techniques. Some of the possible extensions of this effort are:

1) Extension of the technique to multivariate density estimation.

2) More exhaustive analysis of the two sample test should be made to better bound the critical values and power of the test. Theoretical developments in this area may be feasible.

3) Some of the endpoint estimation techniques show promise as tail length discriminators. Additional research along these lines could lead to better methods of tail classification and support definition.

4) Goodness-of-fit tests using the same technique as the two-sample test should be more powerful against some alternatives than existing tests. If used in conjunction with existing tests, they should always increase the power.

5) New techniques of searching the objective space in minimun distance estimation could lead to more effective parameterization of the density. At least a four parameter family is probably necessary to cover unimodal densities. Search time is prohibitively expensive using the scheme presented here to find a global minimum of the distance norm in four parameter space and evaluate it thoroughly.

VI. Bibliography.

1. Abramowitz, Milton and Irene A. Stegun (editors). Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables. New York: Dover Publications, Inc., 1965.

2. Ahmad, I.A. "Nonparametric Estimation of the Location and Scale Parameters Based on Density Estimation," Biometrics, 37: 610 (1981).

3. Ahmad, I.A. "Nonparametric-Estimation of the Location and Scale Parameters Based on Density Estimation," Annals of the Institute of Statistical Mathematics, 34: (1982).

4. Alam, Khurshed. Estimation of a Location Parameter. Technical Report N11. Arlington, Virginia: Office of Naval Research, August, 1971. (AD 736 164).

5. Almquist, Kenneth C. Adaptive Robust Estimation of Population Parameters Using Likelihood Ratio Techniques. MS Thesis, AFIT/GOR/MA/75D-1, Wright-Patterson Air Force Base, Ohio, Air Force Institute of Technology (1975).

6. Anderson, G.L. and R.J.P. DeFigueiredo. "An Adaptive Orthogonal-Series Estimator for Probability Density Functions," Annals of Statistics, 8:347-376 (1980).

7. Anderson, T. W. and D. A. Darling. "Asymptotic Theory of Certain 'Goodness of Fit' Criteria Based on Stochastic Processes," Annals of Mathematical Statistics, 23: 193-212 (1952).

8. Andrews, D.F., et. al. Robust Estimates of Location: Survey and Advances. Princeton, New Jersey: Princeton University Press, 1972.

9. Barlow, R.E., et. al. Statistical Inference Under Order Restrictions. New York: John Wiley and Sons, 1972.

10. Bean, S.J. and C.P. Tsckos. "Developments in Nonparametric Density Estimation," International Statistical Review, 48: 267-287 (1980).

11. Bennett, J.O. Estimation of a Multivariate Probability Density Function Using B-Splines. Doctoral Dissertation. Houston, Texas: Rice University, 1974.

12. Beran, Rudolf. "Minimum Hellinger Distance Estimates for Parametric Models," Annals of Statistics, 5: 455-463 (1977).

13. Bickel, P.J. "On Adaptive Estimation," Annals of Statistics, 10: 647-671 (1982).

14. Bickel, P.J. A Distribution-Free Version of the Smirnov Two Sample Test in the p-Variate Case. Technical Report. Berkley, California: University of California, September, 1967 (AD 695 154).

15. Blom, Gunnar. Statistical Estimates and Transformed Beta Variables. Stockholm: Almquist and Wiksells, 1958.

16. Blum, J. and V. Susarla. "A Fourier Inversion Method for the Estimation of a Density and It's Derivatives," Journal of the Australian Mathematical Society (Series A), 23: 166-171 (1977).

17. Borth, David M. "A Total Entropy Criterion for the Dual Problem of Model Discrimination and Parameter Estimation," Journal of the Royal Statistical Society Series B--Methodological, 37: 77-87 (1975).

18. Brigham, E. Oran. The Fast Fourier Transform. Englewood Cliffs, N.J.: Prentice-Hall, 1974.

19. Brunk, H.D. "On the Range of the Difference Between Hypothetical Distribution Functions and Pyke's Modified Empirical Distribution Function," Annals of Mathematical Statistics, 33: 525-532 (1962).

20. Cacoullos, T. "Estimation of a Multivariate Density," Annals of the Institute of Statistical Mathematics, 18: 179-190 (1965).

21. Campbell, G. "Nonparametric Bivariate Estimation with Randomly Censored Data," Biometrika, 68: 417-422 (1981).

22. Caso, John. Robust Estimation Techniques for Location Parameter Estimation of Symmetric Distributions. MS Thesis, AFIT/GSA/MA/72-3, Wright-Patterson Air Force Base, Ohio, Air Force Institute of Technology (1972).

23. Chan, Lai K. and Lennart S. Rhodin. "Robust Estimation of Location Using Optimally Chosen Sample Quantiles," Technometrics, 22: 225-237 (1980).

24. Cheng, Kuang-Fu. "On Estimation of a Density and It's Derivatives," Presented at IMS 1980 Meeting. Ann Arbor, Michigan, August, 1980.

25. Cheng, K.F. and R.J. Serfling. On Rates of Convergenge in the L2 Norm of Nonparametric Probability Density Estimates. Technical Report. Florida State University at Tallahassee, Department of Statistics, July, 1979 (AD A 072 134).

26. Chung, Kai Lai. A Course in Probability Theory. New York: Academic Press, 1974.

27. Cook, R.D. and M.E. Johnson "A Family of Distributions for Modeling Non-Elliptically Symmetric Multivariate Data," Journal of the Royal Statistical Society, Series B-- Methodological, 43: 210-218 (1981).

28. Cooke, P.J. "Statistical Inference for Bounds of Random Variables," Biometrika, 66: 367-374 (1979).

29. Cooke, P.J. "Optimal Linear Estimation of Bounds of Random Variables," Biometrika, 67: 257-258 (1980).

30. Crain, Bedford R. "An Information Theoretic Approach to Approximating a Probability Distribution," SIAM Journal of Applied Mathematics, 32: 339-346 (March 1977).

31. Cressie, Noel. "Transformations and the Jackknife," Journal of the Royal Statistical Society Series B-- Methodological, 43: 177-182 (1981).

32. Crowder, George E., Jr. Adaptive Estimation Based on a Family of Generalized Exponential Distributions. MS Thesis, AFIT/GOR/MA/77D-2, Wright-Patterson Air Force Base, Ohio, Air Force Institute of Technology (1977).

33. Cuadras, C.M. and J. Auge. "A Continuous General Multivariate Distribution and Its Properties," Communications in Statistics Part A--Theory and Methods, 10: 339-353 (9181).

34. Dahira, Ram C. and John Gurland. A Test of Fit for Bivariate Distributions. Technical Report No. MRC-TSR-1058. Madison, Wisconsin: University of Wisconsin Mathematics Research Center, July, 1971 (AD 729 285).

35. Daniels, Tony G. Robust Estimation of the Generalized t Distribution Using Minimum Distance Estimation. MS Thesis, AFIT/GOR/MA/80D-2, Wright Patterson Air Force Base, Ohio, Air Force Institute of Technology (December 1980).

36. David, F.N. and N.L. Johnson. "The Probability Integral Transform when Parameters are Estimated from the Sample," Biometrika,35: 182-190 (1948).

37. David, H.A. Order Statistics. New York: Wiley, 1970.

38. Deheuvels, P. "An Asymptotic Decomposition for Multivariate Distribution-Free Tests of Independence," Journal of Multivariate Analysis, 11:102-113 (1981).

39. deMontricher, G.F., R.A.Tapia and J.R. Thompson. "Non-Parametric Maximum Liklihood Estimation of Probability Densities by Penalty Function Methods," Annals of Statistics, 3:1329-1348 (1975).

40. Deuser, L.M. and D.G. Lainiotis. Minimum Mean Square Error Approximation of Unknown Probability Distribution Functions. Technical Report. Austin, Texas: University of Texas Electronics Research Center, December 1968 (ad 686 564).

41. Devroye, Luc P. "A Uniform Bound for the Deviation of Empirical Distribution Functions," Journal of Multivariate Analysis, 7: 594-597 (1977).

42. Devroye, Luc P. Nonparametric Discrimination and Density Estimation. Ph.D. Thesis, University of Texas at Austin (1976).

43. Dudewicz, Edward J. and Edward C van der Meulen. Entropy-Based Statistical Inference, I: Testing Hypotheses on Continuous Probability Densities, with Special Reference to Uniformity. Report No. 120. Leuven, Belgium: Department of Mathematics, Katholieke Universiteit Leuven, June 1979.

44. Durbin, J. "Kolmogorov-Smirnov Tests when Parameters are Estimated with Applications to Tests of Exponentiality and Tests on Spacings," Biometrika,62: 5-22 (1975).

45. Dvoretzky, A., J. Kiefer and J. Wolfowitz. "Asymptotic Minimax Character of the Sample Distribution function and of the Classical Multinomial Estimator," Annals of Mathematical Statistics, 27: 642-669 (1956).

46. Easterling, Robert G. "Goodness of Fit and Parameter Estimation," Technometrics, 18: 1-9 (February 1976).

47. Efron, B. The Jackknife, the Bootstrap and Other Resampling Plans. Philadelphia: Society for Industrial and Applied Mathematics, 1982.

48. Efron, B. "Bootstrap Methods: Another Look at the Jackknife," Annals of Statistics, 7: 1-26 (1979).

49. Epanechnikov, V.A. "Nonparametric Estimates of a Multivariate Probability Density," Theory of Probability and It's Applications, 14:153-158 (1969).

50. Fellegi, Ivan P. "Approximate Tests of Independence and Goodness of Fit Based on Stratified Multistage Samples," JASA, 75: 261-268 (1980).

51. Fisher, N.I. "Unbiased Estimation for some Nonparametric Families of Distributions," Annals of Statistics, 10: 603-615 (1982).

52. Fisher, R.A. "On the Mathematical Foundations of Theoretical Statistics," Philosophical Transactions of the Royal Society of London, Series A222: 309-368 (1922).

53. Forth, Charles R. Robust Estimation Techniques for Population Parameters and Regression Coefficients. MS Thesis, AFIT/GOR/MA/74-1, Wright Patterson Air Force Base, Ohio, Air Force Institute of Technology (1974).

54. Foutz, Robert V. "A Test for Goodness-of Fit Based on an Empirical Probability Measure," Annals of Statistics, 8: 989-1001 (1980).

55. Gastwirth, J. "On Robust Procedures," JASA, 61: 929-948 (1966).

56. Gibbons, Jean D. Nonparametric Statistical Inference. New York: McGraw-Hill, 1971.

57. Good, I.J. and R.A. Gaskins. "Non Parametric Roughness Penalties for Probability Densities," Biometrika, 58: 255-277 (1971).

58. Goodman, I.R. Generation of a Multivariate Distribution for Specified Univariate Marginals and Covariance Structure. Technical Report No. NRL-MR-4425. Washington, D.C.: Naval Research Laboratory, May, 1981 (AD A 099 578).

59. Gray, H.L., W.R. Schucany, and T.A. Watkins. "On the Generalized Jackknife and It's Relation to Statistical Differentials," Biometrika, 62: 637-642 (1975).

60. Green, J.R. andY.A.S. Hagazy. "Powerful Modified EDF Goodness-of-Fit Tests," JASA, 71: 204-209 (March 1976).

61. Gruska, Gregory F. "Distributional Analysis of Nonnormal Multivariate Data," ASQC Technical Conference Transactions: 553-560 (1978).

62. Gupta, Shanti S. On Order Statistics and Some Applications of Combinatorial Methods in Statistics. Technical Report. Lafayette, Indiana: Purdue University, 1973 (AD 766 386).

63. Haff, L.R. "Empirical Bayes Estimation of the Multivariate Normal Covariance Matrix," Annals of Statistics, 8: 586-597 (1980).

64. Hall, Peter. "On Estimating the Endpoint of a Distribution," Annals of Statistics, 10: 556-568 (1982).

65. Hall, Peter. "The Influence of Rounding Errors on Some Nonparametric Estimators of a Density and Its Derivatives," SIAM Journal on Applied Mathematics, 42: 390-399 (1982).

66. Hampel, Frank R. "A General Qualitative Definition of Robustness," Annals of Mathematical Statistics, 42: 1887-1896 (1971).

67. Hampel, Frank R. "The Influence Curve and its Role in Robust Estimation," JASA,69: 383-393 (1974).

68. Harp, Tilford. Fully Adaptive Estimation of the Parameters of a t and Half-t Distribution. MS Thesis, AFIT/GOR/MA/79-1, Wright-Patterson Air Force Base, Ohio, Air Force Institute of Technology (1979).

69. Harter, H. Leon. "Another Look at Plotting Positions," Communications in Statistics, Part E - Statistical Reviews. To be published during 1984.

70. Harter, H. Leon. "The Use of Order Statistics in Estimation," Operations Research, 16: 783-798 (1968).

71. Harter, H. Leon, Harry J. Khamis, and Richard E. Lamb. "Modified Kolmogorov-Smirnov Tests of Goodness of Fit," Unpublished Manuscript. Dayton, Ohio: Wright State University, Department of Mathematics and Statistics (1982).

72. Harter, H. Leon, Albert H. Moore and Thomas F. Curry. "Adaptive Robust Estimation of Location and Scale Parameters of Symmetric Populations," Communications in Statistics--Theory and Methods, A8: 1473-1491 (1979).

73. Hartley, H.O. and R.C.Pfaffenberger. On a Family of Lesser Known Goodness of Fit Criteria. Technical Report THEMIS-TR-30. College Station, Texas: Texas A&M University, Institute of Statistics, May, 1971 (AD 724 806).

74. Hazen, Allen. Flood Flows. New York: Wiley, 1930.

75. Heathcote, C.R. "The Integrated Squared Error Estimation of Parameters," Biometrika, 64: 255-264 (1977).

76. Hill, B.M. "A Simple General Approach to Inference About the Tail of a Distribution," Annals of Statistics, 3: 1163-1174 (1975).

77. Hill, D.L. and P.V. Rao. "Tests of Symmetry Based on Cramer von Mises Statistics," Biometrika, 64: 489-494 (1977).

78. Hodges, J.L. and E.L. Lehmann. "Estimates of Location Based on Rank Tests," Annals of Mathematical Statistics, 34: 598-611 (1963).

79. Hogg, Robert V. "Adaptive Robust Procedures: A Partial Review and Some Suggestions for Future Applications and Theory," JASA, 69: 909-927 (1974).

80. Holcomb, R.L., R.A. Kronmal and M.E. Tartar. "A Description of New Computer Methods for Estimating the Population Density," Proceedings from the Association of Computing Machinery, 22. New York: Thompson Book Co., 511-519 (1967).

81. Hsu, Jason C. "Robust and Nonparametric Subset Selection Procedures," Communications in Statistics, Part A--Theory and Methods, A9: 1439-1459 (1980).

82. Huber, Peter J. "Robust Location of a Location Parameter," Annals of Mathematical Statistics, 35: 73-101 (1964).

83. Huber, Peter J. "The 1972 Wald Lecture: Robust Statistics: A Review," Annals of Mathematical Statistics, 43: 1041-1067 (1972).

84. James, William L. Minimum Distance Estimation Techniques Based on a Family of Gamma Distributions Using Robust Estimation and Monte Carlo Simulation. MS Thesis, AFIT/ GOR/MA/80d-3, Wright-Patterson Air Force Base, Ohio, Air Force Institute of Technology (1980).

85. Jones, D.H. "Efficient Adaptive Distribution-Free Test for Location," JASA, 74: 822-828 (1979).

86. Jorgensen, Loren W. Robust Estimation of Location and Scale Parameters. MS Thesis, AFIT/GSA/MA/73-2, Wright-Patterson Air Force Base, Ohio, Air Force Institute of Technology (1973).

87. Kaplan, E.L. and P.Meier. "Nonparametric Estimation from Incomplete Observations," JASA, 53: 457-481 (1958).

88. Kapur, R.C. and L.R. Lamberson. Reliability in Engineering Design. New York: Wiley, 1977.

89. Kiefer, J. "On Large Deviations of the Empirical Distribution Function of Vector Chance Variables," Annals of Mathematical Statistics, 32: 649-660 (1961).

90. Kiefer, J. and J. Wolfowitz. "On the Deviations of the Empiric Distribution Function of Vector Chance Variables," Transactions of the American Mathematical Society, 87: 173-186 (1958).

91. Kim, B.K. and J. Van Ryzin. "Uniform Consistency of a Histogram Density Estimator and Modal Estimation," Communications in Statistics, 4: 303-315 (1975).

92. Kimball, B.F. "On the Choice of Plotting Positions on Probability Paper," JASA, 55: 546-560 (1960).

93. Kingman, J.F.C., and S.J.Taylor. Introduction to Measure and Probability. Cambridge, England: Cambridge University Press, 1966.

94. Klonias, V.K. "Consistency of Two Maximum Penalized Liklihood Estimates of the Probability Density Function," Annals of Statistics, 10: 811-824 (1982).

95. Kochar, Subhash C. "Distribution Free Comparison of Two Probability Distributions with Reference to Their Hazard Rates," Biometrika, 66: 437-442 (1979).

96. Kochar, Subhash C. "A New Distribution Test for the Equality of Two Failure Rates," Biometrika, 68: 423-426 (1981).

97. Konakov, V.D. "Some Problems in Nonparametric Estimation of a Probability Density," Theory of Probability and Its Applications, 25: 638-639 (1981).

98. Konakov, V.D. "Complete Asymptotic Expansions for Maximum Deviation of an Empirical Density Function," Theory of Probability and Its Applications, 22: 632-634 (1977).

99. Kotz, Samuel. Multivariate Distributions at a Crossroad. Technical Report. Philadelphia, Pennsylvania: Temple University, Department of Mathematics, July, 1974 (AD A 000 391).

100. Kotz, Samuel. Multivariate Statistical Models: Abstracted Subjected-Classified Bibliography. Technical Report. Philadelphia, Pennsylvania: Temple University, Department of Mathematics, 1974 (AD A 000 390).

101. Kotz, Samuel. Annotated and Abstracted Bibliography on Multivariate Statistical ModelsTechnical Report. Philadelphia, Pennsylvania: Temple University, Department of Mathematics, 1974 (AD A 000 389).

102. Kowar, Ramesh M. and Myles Hollander. Empirical Bayes Estimation of a Distribution Function. Tech. Rpt. FSU-Statistics-M288. Tallahassee, Florida: Florida State University, Dept. of Statistics, March, 1974 (AD 778 455).

103. Koziol, J.A. "Test for Bivariate Symmetry Based on the Empirical Distribution Function," Communications in Statistics--Theory and Methods, 8: 207-221 (1979).

104. Koziol, J.A. "Goodness of Fit Tests Based on Empirical Distribution Function for Uniform Spacings," Journal of the Royal Statistical Society, Series B--Methodological, 39: 333-336 (1977).

105. Kozoil, J.A. and S.B.Green. "A Cramer-von Mises Statistic for Randomly Censored Data," Biometrika, 63: 465-473 (1976).

106. Kreimerman, Joseph. A Bivariate Test of Goodness of Fit based on a Gradually Increasing Number of Order Statistics. Technical Report No. TR-250. Ithaca, N.Y. Cornell University, Department of Operations Research, March, 1975 (AD A 008 205).

107. Kronmal, R.A. and M.E. Tartar. "The Estimation of Probability Densities and Cumulatives by Fourier Series Methods," JASA, 63: 925-952 (1968).

108. Lamperti, John. Probability. New York: W.A. Benjamin, Inc, 1966.

109. Launer, Robert L. and Graham N. Wilkinson. Robustness in Statistics. New York: Academic Press, 1979.

110. Leonard, T. The Empirical Bayesian Estimation of a One-Dimensional Function. Technical Report, University of Wisconsin-Madison, Mathematics Research Center (1982).

111. Leonard, T. Bayes Estimation of a Multivariate Density. Technical Report. University of Wisconsin-Madison, Mathematics Research Center, February, 1982. (AD A 114 579).

112. Lii, Keh-Shin and M.Rosenblatt. "Asymptotic Behavior of a Spline Estimate of a Density Function," Computation and Mathematics with Applications, 1: 223-235 (1975).

113. Lilliefors,Hubert W. "On the Kolmogorov-Smirnov Test for Normality with Mean and Variance Unknown," JASA, 399-402 (1967).

114. Littell, R.C., J.T. McClave, and W.W. Offen. "Goodness of Fit Tests for the Two Parameter Weibull Distribution," Communications in Statistics--Simula. Computa., B8: 257-269 (1979).

115. Mack, Y.P. and M. Rosenblatt. "Multivariate K-Nearest Neighbor Density Estimates," Journal of Multivariate Analysis, 9: 1-15 (1979).

116. MacQueen, James and Jacob Marschak. "Partial Knowledge, Entropy and Estimation," Proceedings of the National Academy of Sciences, 3819-3824 (October 1975).

117. Mann, H.B. and D.R. Whitney. "On a Test of Whether One of Two Random Variables is Stochastically Larger than the Other," Annals of Mathematical Statistics, 18: 50-60 (1947).

118. Mann, N.R., E.M.Scheuer and K.W. Fertig. "A New Goodness of Fit Test for the Two Parameter Weibull or Extreme Value Distribution with Unknown Parameters," Communications in Statistics, 2: 383-400 (1973).

119. McGrath, E.J. and D.C. Irving. Techniques for Efficient Monte Carlo Simulation. Volume II. Random Number Generation for Selected Probability Distributions. SAI Report SAI-72-590-1, Arlington, Virginia: Office of Naval Research, March 1973 (AD 762 722).

120. McNeese, Larry B. Adaptive Minimum Distance Estimation Techniques Based on a Family of Generalized Exponential Power Distributions. MS Thesis, AFIT/GOR/MA/80d, Wright-Patterson Air Force Base, Ohio, Air Force Institute of Technology (December 1980).

121. McNichols, D.T. and W.J.Padgett. Kernel Density Estimation Under Random Censorship. Technical Report No. 74. Columbia, S.C.: University of South Carolina, Department of Mathematics and Statistics, October, 1981.

122. Michael, John R. and William R. Schucany. The Influence Curve and Goodness of Fit. Technical Report TR-137. Dallas, Texas: Southern Methodist University, Department of Statistics, May, 1980 (AD A 085 593).

123. Mihalko, Daniel P. and David S. Moore. "Chi-Square Tests of Fit for Type II Censored Data," Annals of Statistics, 8: 625-644 (May 1980).

124. Miller, James E., Jr. Continuous Density Approximation on a Bounded Interval Using Information Theoretic Concepts. Ph.D. Dissertation, AFIT/DS/MA/80-1, Wright-Patterson Air Force Base, Ohio, Air Force Institute of Technology (1980).

125. Miller, Rupert G. "The Jackknife-- A Review," Biometrika, 61: 1-15 (1974).

126. Moore, Albert H. "Extension of Monte Carlo Techniques for Obtaining System Reliability Confidence Limits from Component Test Data," Proceedings of the National Aerospace Electronics Conference, 459-463 (may 1965).

127. Moore, Albert H. Robust Statistical Inference. Notes from a short course presented at the Air Force Institute of Technology, Wright-Patterson Air Force Base, Ohio, February 1981.

128. Moore, D.S. and E.G. Henrichon. "Uniform Consistency of some Estimates of a Density Function," Annals of Mathematical Statistics, 40: 1499-1502 (1969).

129. Morgera, S.D. "Structured Estimation, .2. Multivariate Probability Density Estimation," IEEE Transactions on Information Theory, 27: 607-622 (1981).

130. Nadaraya, E.A. "Some Problems in Nonparametric Estimation of Probability Densities and a Regression Curve," Theory of Probability and Its Applications, 25: 637-638 (1981).

131. Nadaraya, E.A. "On Nonparametric Estimates of Density Functions and Regression Curves," Theory of Probability and Its Applications, 10: 186-190 (1965).

132. Nadaraya, E.A. "Remarks on Nonparametric Estimates for Density Functions and Regression Curves," Theory of Probability and Its Applications, 15: 134-137 (1970).

133. O'Reilly, Federico J. and Michael A. Stephens. Characterizations and Goodness of Fit Tests. Technical Report TR-302. Stanford University, Department of Statistics, June, 1981 (AD A 102 167).

134. Padgett, W.J., R.L. Taylor and L.J. Wei. Nonparametric Bayes Estimation of Distribution Functions and the Study of Probability Density Estimates. Technical Report. Columbia, S.C.: South Carolina University, Department of Mathematics and Statistics, June, 1981 (AD A 102 413).

135. Parr, William C. Minimum Distance and Robust Estimation. Ph.D. Dissertation, Dallas, Texas, Southern Methodist University, 1978.

136. Parr, William C. Minimum Distance Estimation: A Bibliography. Unpublished Manuscript. Institute of Statistics, Texas A&M University, College Station, Texas, 1980.

137. Parr, William C. and William R. Schucany. "The Jackknife: A Bibliography," International Statistical Review, 48: 73-78 (1980).

138. Parr, William C. and T. Dewet. On Minimum CVM-Norm Parameter Estimation. Unpublished Manuscript. Institute of Statistics, Texas A&M University, College Station, Texas, and Department of Mathematical Statistics, Rhodes University, Grahamstown, South Africa, 1979.

139. Parzen, Emanuel. "On the Estimation of a Probability Density Function and the Mode," Annals of Mathematical Statistics, 33:1065-1076 (1962).

140. Parzen, Emanuel. Nonparametric Statistical Data Science: A Unified Approach Based on Density Estimation and Testing for White Noise. Tech. Rpt. Univ. of New York, Amherst, Jan. 1977. (AD A 051 090).

141. Parzen, Emanuel. "Nonparametric Statistical Data Modeling," JASA, 74: 105-121 (1979).

142. Patil, S.A., J.L. Kovner and K.P. Burnham. "Optimum Nonparametric Estimation of Population Density Based on Ordered Distances," Biometriks, 38: 243-248 (1982).

143. Pearson, Karl. "Contributions to the Mathematical Theory of Evolution. II.Skew Variations in Homogeneous Material," Philosophical Transactions of the Royal Society of London, Series A 186: 343-414 (1895).

144. Pennington, Ralph H. Introductory Computer Methods and Numerical Analysis. New York: Macmillan, 1965.

145. Penrod, C.S. Nonparametric Estimation with Local Rules. Ph.D. Dissertation, University of Texas at Austin 1976.

146. Pettitt, A.N. "Testing for Bivariate Normality Using the Empirical Distribution Function," Communications in Statistics--Theory and Methods, 8: 699-712 (1979).

147. Phadia, Eswarlal G. "A Note on Empirical Bayes Estimation of a Distribution Function Based on Censored Data," Annals of Statistics, 8: 226-229 (1980).

148. Phadia, Eswarlal G. On Estimation of a Cumulative Distribution Function. Ph.D. Dissertation, Columbus, Ohio, Ohio State University, 1971.

149. Pollard, D. "The Minimum Distance Method of Testing," Metrika, 27: 43-70 (1980).

150. Puri, Madan Lal. Nonparametric Techniques in Statistical Inference. Technical Report. Bloomington, Indiana: Indiana University, Department of Statistics, 1970 (AD 720 284).

151. Puri, Madan L. and Lanh T. Tran. "Empirical Distribution Functions and Functions of Order Statistics for Mixing Random Variables," Journal of Multivariate Analysis, 10: 405-425 (1980).

152. Pyke, Ronald. "The Supremum and Infinum of the Poisson Process," Annals of Mathematical Statistics, 30: 568-576 (1959).

153. Pyke, Ronald. "Spacings," Journal of the Royal Statistical Society, Series B, 27: 395-436 (1965).

154. Quenouille, M.H. "Approximate Tests of Correlation in Time-Series," Biometrika, 43:353-360 (1956).

155. Ramberg, John S., et. al. "A Probability Distribution and its Uses in Fitting Data," Technometrics, 21: 201-214 (1979).

156. Rao, C. Radhakrishna. Linear Statistical Inference and Its Applications. New York: Wiley, 1965.

157. Reiss, R.D. "Nonparametric Estimation of Smooth Distribution Functions," Scandinavian Journal of Statistics, 89:116-119 (1982).

158. Reiss, R.D. "On Minimum Distance Estimators for Unimodal Densities," Metrika, 23: 7-14 (1976).

159. Rosenblatt, M. "A Quadratic Measure of Deviation of Two-Dimensional Density Estimates and a Test of Independence," Annals of Statistics, 3: 1-14 (1975).

160. Rosenblatt, Murray. "Remarks on Some Nonparametric Estimates of a Density Function," Annals of Mathematical Statistics, 27: 832-837 (1956).

161. Rothman, E.D. and M. Woodroofe. "A Cramer-von Mises Type Statistic for Testing Symmetry," Annals of Mathematical Statistics, 43: 2035-2038 (1972).

162. Rudemo, M. "Empirical Choice of Histograms and Kernel Density Estimators," Scandinavian Journal of Statistics, 9: 65-78 (1982).

163. Rugg, Bernard J. Adaptive Robust Estimation of Location and Scale Parameters Using Selected Discriminants. MS Thesis, AFIT/GOR/MA/74D-3, Wright-Patterson Air Force Base, Ohio, Air Force Institute of Technology (1974).

164. Rustagi, J.S. and S. Dynin. Jackknifing Kernel Type Density Estimators. Technical Report No. 280. Department of Statistics, Ohio State University, Columbus, Ohio (1983).

165. Sahler, W. "A Survey on Distribution-Free Statistics Based on Distances Between Distribution Functions," Metrika, 13: 149-169 (1968).

166. Sahler, W. "Estimation By Minimum Discrepancy Methods," Metrika, 15: 85-106 (1970).

167. Saniga, Erwin M. and James A. Miles. "Power of Some Standard Goodness-of-Fit Tests of Normality Against Asymmetric Stable Alternatives," JASA, 74: 861-865 (1979).

168. Saunders, Roy and Purushottam Laud. "The Multidimensional Kolmogorov Goodness-of-Fit Test," Biometrika, 67: 237 (1980).

169. Saxena, A.K. "Complex Multivariate Statistical Analysis: Annotated Bibliography," International Statistical Review, 46: 209-214 (1978).

170. Schoenberg, I.J. "Notes on Spline Functions II: On the Smoothing of Histograms," Technical Report No. 1222, Univ. of Wis. at Madison, Mathematics Research Center (1972)

171. Schoenberg, I.J. Splines and Histograms. Technical Report No. 1273, University of Wisconsin at Madison, Mathematics Research Center (1972)

172. Schreiber, F. "Generalized Equations for the Objective Empirical Distribution Function," AEU-Archiv fur Elektronik und Ubertragungstechnik - Electronics and Communication, 36: 168-172 (1982).

173. Schucany, W.R. and J.P.Sommers. "Improvements of Kernel Type Density Estimators," JASA, 72: 420-423 (1977).

174. Schuster, Eugene F. "Estimation of a Probability Density Function and Its Derivatives," Annals of Mathematical Statistics, 40: 1187-1195 (1969).

175. Schuster, Eugene F. "On the Goodness-of-Fit Problem for Continuous Symmetric Distributions," JASA, 68: 713-715 (1973). Corrigenda JASA, 69: 288 (1974).

176. Schuster, Eugene F. "Estimating the Distribution Function of a Symmetric Distribution," Biometrika, 62: 631-636 (1975).

177. Schwartz, Stewart. "Estimation of a Probability Density by an Orthogonal Series," Annals of Mathematical Statistics, 38: 1261-1265 (1967).

178. Scott, D.W., R.A. Tapia and J.R. Thompson. "Nonparametric Density Estimation by Discrete Maximum Penalized-Likelihood Criteria," Annals of Statistics, 8: 820-832 (1980).

179. Scott, D.W., R.A. Tapia and J.R. Thompson. "Kernel Density Estimation Revisited," Nonlinear Analysis, 1: 339-372 (1977).

180. Scott, D.W., R.A. Tapia and J.R. Thompson. Sen, Pranab Kumar. "Nonparametric Tests for Multivariate Interchangeability. Part 1: Problems of Location and Scale in Bivariate Distributions," Sankhya: The Indian Journal of Statistics, Series A, 29: (1967).

181. Shorack, G.R. and J.A. Wellner. "Linear Bounds on Empirical Distribution Function," Annals of Probability, 6: 349-353 (1978).

182. Sievers, Gerald L. and John Kapenga. Approximate Empirical Distributions for the Computation of Nonparametric Statistics. Technical Report TR-64. Kalamazoo, Michigan: Western Michigan University, Department of Mathematics, February, 1981 (AD A 100 223).

183. Silverman, B.W. "On the Estimation of a Probability Density Function by the Maximum Penalized Liklihood Method," Annals of Statistics, 10: 795-810 (1982).

184. Silverman, B.W. "Choosing the Window Width when Estimating a Density," Biometrika, 65: 1-13 (1978).

185. Simons, Gordon. Generalized Cumulative Distribution Functions: I. The Linear Case with Applications to Nonparametric Statistics. Technical Report. Chapel Hill, North Carolina: University of North Carolina, Department of Statistics, August, 1972 (AD 751 286).

186. Singh, Jagbir. "Minimum Variance Unbiased Estimation of Probability Densities," Australian Journal of Statistics, 22: 328-331 (1980).

187. Singh, R.S. "Mean Square Error of Estimates of a Density and its Derivatives," Biometrika, 66:177-180 (1979).

188. Sinha, Bimal K. and H.S. Wieand. "Bounds on the Efficiencies of Four Commonly Used Nonparametric Tests of Location," Sankhya, Series B, Indian Journal of Statistics, 39: 121-129 (1977).

189. Smaga, Edward. "Smooth Empirical Distribution Function," Przeglad Statystyczny, 25.1: Warsaw, Poland (1978).

190. Smirnov, N.V. "Estimate of Deviation Between Empirical Distribution Functions in Two Independent Samples," (in Russian) Bulletin of Moscow University, 2: 3-16 (1939).

191. Smith, R. M. and L.J. Bain. "Correlation Type Goodness-of-Fit Statistics with Censored Samples," Communications in Statistics--Theory and Methods, A5: 119-132 (1976).

192. Srikanthan, R. and T.A. mcMahon. "Log Pearson III Distribution - An Empirically Derived Plotting Position," Journal of Hydrology, 52: 161-163 (1981).

193. Stephens, M.A. "Use of Kolmogorov-Smirnov, Cramer-von Mises and Relaxed Statistics Without Extensive Tables," Journal of the Royal Statistical Society, Series B, 32, No. 1:115-122 (1970).

194. Stephens, M.A. "EDF Statistics for Goodness-of-Fit and Some Comparisons," JASA, 69: 730-737 (1974).

195. Stephens, M.A. "Goodness-of-Fit for the Extreme Value Distribution," Biometrika, 64: 583-588 (1977).

196. Stephens, M.A. EDF Statistics for Goodness-of-Fit. Part 2. Power Studies. Part 3. Miscellaneous Complements. Tech. Report. Palo Alto, California: Stanford University, Department of Statistics, December, 1972 (AD 758 670).

197. Stigler, Stephen M. Simon Newcomb, Percy Daniell, and the History of Robust Estimation, 1385-1920. Technical Report No. 319. Arlington, Virginia: Office of Naval Research, December 1972 (AD 757 026).

198. Stigler, Stephen M. "Do Robust Statistics Work with Real Data?" (With Discussants), Annals of Statistics, 5:1055-1098 (1977).

199. Stigler, Stephen M. "Studies in the History of Probability and Statistics XXXVIII--R.H. Smith, A Victorian Interest in Robustness," Biometrika, 67: 217-221 (1980).

200. Stone, Charles J. "Optimal Rates of Convergenge for Nonparametric Estimators," Annals of Statistics, 8: 1348-1360 (1980).

201. Susarla, V. and G. Walter. "Estimation of a Multivariate Density Function Using Delta Sequences," Annals of Statistics, 9: 347-355 (1981).

202. Sweeder, J. Nonparametric Estimation of Distribution and Density Functions With Applications. PhD Dissertation. AFIT/DS/MA/82-1. Wright-Patterson Air Force Base, Ohio: Air Force Institute of Technology, 1982

203. Tapia, Richard A. and James R. Thompson. Nonparametric Probability Density Estimation. Baltimore: Johns Hopkins University Press, 1978.

204. Tsokos, Chris P. and A. Rust III. "Recent Developments in Nonparametric Estimation of Probability," Applied Stochastic Processes: 269-281 (1980).

205. Tukey, J.W. "Bias and Confidence in Not-quite Large Samples," Annals of Mathematical Statistics, 29: 614 (1958).

206. Turnbull, Bruce W. "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," Journal of the Royal Statistical Society Series B--Methodological, 38: 25 (1976).

207. Vanzuijlen, M.C.A. "Properties of the Empirical Distribution Function for Independent Non-Identically Distributed Random Vectors," Annals of Probability, 10: 108-123 (1982).

208. Vogt, Herbert. "Concerning a Variant of the Empirical Distribution Function," Metrika, 25: 49-58 (1978).

209. Wagner, T.J. The Study of Distribution-Free Performance Bounds for Nonparametric Discrimination Algorithms. Technical Report. University of Texas at Austin, Department of Electrical Engineering, June, 1977 (AD A 042 685).

210. Wagner, T.J. "Nonparametric Estimates of Probability Densities," IEEE Transactions on Information Theory, IT-21: 438-440 (1975).

211. Wagner, T.J. "Strong Consistency of a Nonparametric Estimate of a Density Function," IEEE Transactions on Systems, Man, and Cybernetics, 3: 289-290 (1973).

212. Wagner, T.J. and W.H. Rogers. "A Finite Sample Distribution-Free Performance Bound for Local Discrimination Rules," Accepted by Annals of Statistics

213. Wagner, T.J. and C.S. Penrod. "Risk Estimation for Nonparametric Discrimination and Estimation Rules: A Simulation Study," Submitted to IEEE Transactions on Systems, Man, and Cybernetics.

214. Wahba, Grace. "Optimal Convergence Properties of Variable Knot, Kernel, and Orthogonal Series Methods for Density Estimation," Annals of Statistics, 3: 15-29 (1975).

215. Wahba, Grace. Interpolating Spline Methods for Density Estimation. II.Variable Knots. Technical Report 337. University of Wisconsin at Madison, Department of Statistics, 1973.

216. Wahba, Grace "A Polynomial Algorithm for Density Estimation," Annals of Mathematical Statistics, 42: 1870-1886 (1971).

217. Wahba, Grace and A. Wold. "Periodic Splines for Spectral Density Estimation: The Use of Cross Validation for Determining the Degree of Smoothing," Communication Statistics, 4: 125-141 (1975).

218. Wald, A. and J. Wolfowitz. "On a Test of Whether Two Samples Are From the Same Population," Annals of Mathematical Statistics, 11: 147-162 (1940).

219. Walter, Gilbert G. "Properties of Hermite Series Estimation of Probability Density," Annals of Statistics, 5: 1258-1264 (1977).

220. Walter, Gilbert G. and J. Blum. "Probability Density Estimation Using Delta Sequences," Annals of Statistics, 7: 328-340 (1979).

221. Waterman, M.S. and D.E. Whiteman. "Estimation of Probability Densities by Empirical Density Functions," International Journal of Mathematical Education in Science and Technology, 9: 127-137 (1978).

222. Watson, G.S. "Density Estimation by Orthogonal Series," Annals of Mathematical Statistics, 34: 1496-1498 (1969).

223. Watson, G.S. and M.R. Leadbetter. "Hazard Analysis II," Sankhya, 26A: 101-116 (1964).

224. Watson, G.S. and M.R. Leadbetter. "On the Estimation of the Probability Density I," Annals of Mathematical Statistics, 34: 480-491 (1963).

225. Wegman, Edward J. "Nonparametric Probability Density Estimation: I. A Summary of Available Methods," Technometrics, 14: 533-546 (1972).

226. Wegman, Edward J. "Nonparametric Probability Density Estimation: II. A Comparison of Density Estimation Methods," Journal of Statistical Computations and Simulation, 1: 225-245 (1972).

227. Wegman, Edward J. and H.I. Davies. "Remarks on Some Recursive Estimators of a Probability Density," Annals of Statistics, 7: 316-327 (1979).

228. Weibull, Waloddi. Outline of a Theory of Powerful Selection of Distribution Functions. Technical Report Scientific-C. Lausanne, Switzerland: March, 1971 (AD 725 037).

229. Weiss, L. and J. Wolfowitz. "Asymptotically Efficient Nonparametric Estimators of Location and Scale Parameters," Z. Wahrscheinlichkeits - Theorie Revw Gebiete, 16: 134-150 (1970).

230. Westergaard, Harald. Contributions to the History of Statistics. New York, Agathon, 1968.

231. White, John S. "The Moments of Log-Weibull Order Statistics," Technometrics, 11: 373-386 (1969).

232. Whittle, P. "On Smoothing of Probability Density Functions," Journal of the Royal Statistical Society, Series B, 20: 334-343 (1958).

233. Wold, S. "Spline Functions in Data Analysis," Technometrics, 16: 1-11 (1974).

234. Wolfowitz, J. "The Minimum Distance Method," Annals of Mathematical Statistics, 28: 75-88 (1957).

235. Wolfowitz, J. "Convergence of the Empiric Distribution Function on Half-spaces," Contributions in Probability and Statistics. Edited by I. Olkin, et.al., Stanford University Press, California, 504-507 (1960).

236. Woodroofe, Michael. "On Choosing a Delta Sequence," Annals of Mathematical Statistics, 41: 1665-1671 (1970).

237. Wright, Ian W. Spline Methods in Statistics. Technical Report No. 77-1307. Bolling Air Force Base, D.C., Air Force Office of Scientific Research, 1977 (AD A 049 197).

238. Yu, George C.S. "Power Bounds on Some Nonparametric Test Procedures for Censored Data," Sankhya, Series B, Indian Journal of Statistics, 39: 279-283 (1977).

239. Zacks, Shelemyahu. The Theory of Statistical Inference. New York: Wiley, 1971.

VITA

Major Ronald P. Fuchs was born on 1 August 1944 in Providence, Rhode Island. He graduated from high school in Arlington, Virginia in 1962 and attended Virginia Polytechnic Institute and State University (VPI&SU) as a cooperative education student and received a degree of Bachelor of Science in Aerospace Engineering in June 1967. He was commissioned in the USAF through ROTC. He continued at VPI&SU until he received his Master of Science in Control System Engineering in 1968.

Major Fuchs has served as Chairman of the Government Fluidics Coordinating Group (1971), and as a member of the AIAA National Technical Committee on Guidance and Control (1974-76). He has numerous technical publications in diverse fields.

He served in the Air Force as a project manager on the Space Defense System at Space Division; as Assistant Professor of Astronautics and Director of the Guidance and Control Laboratory at the USAF Academy; and as project manager for the F-16 program at Aeronautical Systems Division. Major Fuchs entered the School of Engineering, Air Force Institute of Technology, in June 1981.

Permanent Address: 3885 Tusico Pl.
Fairfax, VA 22030

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a. SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for Public Release; Distribution Unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S)	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION School of Engineering AF Institute of Technology 6b. OFFICE SYMBOL (If applicable) AFIT/EN	7a. NAME OF MONITORING ORGANIZATION
6c. ADDRESS (City, State and ZIP Code) Wright-Patterson AFB, OH, 45433	7b. ADDRESS (City, State and ZIP Code)
8a. NAME OF FUNDING/SPONSORING ORGANIZATION F-16 SPO Aero. Systems Division 8b. OFFICE SYMBOL (If applicable) ASD/YPP	9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
8c. ADDRESS (City, State and ZIP Code) Wright-Patterson AFB, OH, 45433	10. SOURCE OF FUNDING NOS. PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO. / WORK UNIT NO.
11. TITLE (Include Security Classification) See Box 19	
12. PERSONAL AUTHOR(S) Ronald P. Fuchs	
13a. TYPE OF REPORT PhD Dissertation 13b. TIME COVERED FROM _____ TO _____	14. DATE OF REPORT (Yr., Mo., Day) 84 May 15. PAGE COUNT 148
16. SUPPLEMENTARY NOTATION	

17. COSATI CODES			18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB. GR.	
12	01		Statistical Functions, Statistical Tests, Non-parametric Statistics, Probability Density Functions, Statistics

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Title: A NON-PARAMETRIC PROBABILITY DENSITY ESTIMATOR AND SOME APPLICATIONS

Chairman: Albert H. Moore

Approved for public release: IAW AFR 190-17.

LYNN E. WOLAVER 25 Feb 81

Dean for Research and Professional Development

Air Force Institute of Technology (AIC)

Wright-Patterson AFB OH 45432

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT UNCLASSIFIED/UNLIMITED XX SAME AS RPT. ☐ DTIC USERS ☐	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Albert H. Moore, Professor	22b. TELEPHONE NUMBER (Include Area Code) 513-255-3098 22c. OFFICE SYMBOL AFIT/ENC

DD FORM 1473, 83 APR EDITION OF 1 JAN 73 IS OBSOLETE.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE

A new non-parametric probability density estimator is developed which has the following properties:

1) It yields a continuous, non-negative and piecewise linear estimate of a probability density function.

2) It converges to the true density function if the true density has no more than a finite number of discontinuities of a form where the value of the function at the discontinuity can be considered the average of the limiting values on either side of the discontinuity.

3) It requires no user supplied parameters.

The estimator is shown to have significantly better error properties, for certain classes of distributions, than existing density estimators. The quality of the estimate is discussed, tabulated and graphically demonstrated. Applications, including parameterization, small sample analysis, and two sample tests are presented. These newly developed applications are shown to improve upon the generally accepted existing techniques. Guidelines for choosing a density estimation method along with an organized approach to method selection are discussed.

END

FILMED

4-85

DTIC